New techniques in geometric and discrete optimization

Hurkens, Cornelius Antonius Josephus

Publication date:
1989

Document Version
Publisher's PDF, also known as Version of record

New Techniques in Geometric and Discrete Optimization

Cor Hurkens

New Techniques in Geometric and Discrete Optimization

New Techniques in Geometric and Discrete Optimization

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Brabant,
op gezag van de rector magnificus,
prof. dr. R.A. de Moor, in het
openbaar te verdedigen ten overstaan
van een door het college van dekanen
aangewezen commissie in de aula van de
Universiteit op donderdag 12 oktober 1989
te 16.15 uur

door

Cornelius Antonius Josephus Hurkens

geboren te Haps

Promotor: Prof. Dr. A. Schrijver.

PREFACE.

This thesis is a collection of seven papers on problems of a geometric and discrete nature, each motivated by questions arising in combinatorial optimization, integer programming and discrete geometry. The emphasis lies on the derivation of sharp bounds and reductions. Three papers have appeared as articles, one has appeared as a contribution to a proceedings, while two others have been accepted for publication. The last paper has not yet been submitted to a scientific journal.

The first paper, '**On fractional multicommodity flows and distance functions**', is joint work with A. Schrijver and É. Tardos. It has appeared in *Discrete Mathematics* 73 (1988/89), 99-109, and discusses some results on the existence of (integral) multicommodity flows in planar graphs. In particular the issue is raised under what conditions the existence of a fractional solution to a multicommodity flow problem implies the existence of an integral solution.

The second paper, '**On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems**', is again co-authored work with A. Schrijver. It appeared in *SIAM Journal on Discrete Mathematics* vol. 2, no. 1, (February 1989) 68-72, and presents bounds on the relative size of a 'maximal' family of pairwise disjoint sets among a given family of sets of size k. The research initiated from the study of the behaviour of certain heuristics for packing problems proposed by J.K. Lenstra. Part of the research for this paper was performed while the authors were visiting the Rutgers Center for Operations Research, Rutgers University, New Brunswick, New Jersey.

The third and fourth paper are strongly related. The third paper, '**On the diameter of the edge cover polytope**', characterizes the adjacency of edge covers and provides a bound on the distance between two edge covers. As a result the diameter of the edge cover polytope is determined. The paper will appear in *Journal of Combinatorial Theory (Series B)*.

In an effort to get a more elegant proof of the main theorem involved (viz. on the distance between two edge covers) a generalization was found, leading to results described in the fourth paper, **'On the diameter of the *b*-matching polytope'**. In this paper it is shown that the diameter of the *b*-matching polytope is equal to the cardinality of the largest *b*-matching. It appeared in the proceedings of the Seventh Hungarian Conference on Combinatorics, Finite and Infinite Sets, held in Eger, 1987 (*Combinatorics* (A. Hajnal, L. Lovász and V.T. Sós, eds.), North-Holland, Amsterdam, 1988, pp. 301-307).

In the fifth paper, **'Blowing up Convex Sets in the Plane'**, a problem raised by R. Kannan and L. Lovász is solved. This problem in Euclidean geometry arose in their study of the theory of lattices and the geometry of numbers. The paper will appear in *Linear Algebra and its Applications*. The result described in this paper is the following: for a convex body K in $\mathbb{R}^2$ and a lattice $\mathcal{L}$, with the property that every line in $\mathbb{R}^2$ intersects $K + \mathcal{L}$, we have that $\alpha \cdot K + \mathcal{L}$ covers $\mathbb{R}^2$, if $\alpha \geq 1 + 2/\sqrt{3}$. Furthermore, this bound is best possible.

The last two papers are again related. They both start from a solution of A. Schrijver to the problem of finding vertex-disjoint circuits of prescribed homotopies in a graph embedded on a compact surface. Problems of this type are motivated by the design of VLSI-circuits, and by the big Graph Minors project of N. Robertson and P.D. Seymour. An early version of Schrijver's algorithm contains as a sub-routine the solution of a certain system of linear inequalities in integers. In general, solving linear inequalities in integers is a 'hard' problem. Due to the special form of the constraint matrix involved however it is possible to solve this problem 'easily'. Regarding the matrix as the adjacency matrix of a bidirected graph it is possible to formulate necessary and sufficient conditions for the existence of an integral solution to the inequality system at hand in terms of conditions on cycles in the bidirected graph.

In the sixth paper, **'On the existence of an integral potential in a weighted bidirected graph'**, it is shown that it suffices to enforce these conditions on a reduced set of rather simple cycles. This paper has appeared in *Linear Algebra and its Applications* **114/115** (1989) 541-553.

Actually Schrijver's theorem gives necessary and sufficient conditions for the existence of vertex-disjoint circuits of prescribed homotopies in a graph embedded on a compact surface in terms of conditions on closed curves on the surface. In the seventh paper, **'Reduction of cut-conditions on compact surfaces'**, we give a partial reduction of the set of necessary and sufficient conditions at hand. This paper has not yet been submitted to a scientific journal. It is the first step in the transformation of Schrijver's theorem to a so-called 'good characterization'.

ACKNOWLEDGMENTS.

The results in this thesis are part of the outcome of a fruitful period of scientific research starting from September 1, 1985, in which I participated in the project "Polyhedral and Polynomial Methods in Combinatorial Optimization" under supervision of prof. dr. Alexander Schrijver. The research was financially supported by Z.W.O., the Netherlands Organisation for the Advancement of Pure Research, via the 'Stichting Mathematisch Centrum'. I thank these organisations for their generous support throughout the years.

Under the stimulating and exemplary guidance of Lex Schrijver I have learnt a lot about parts of mathematics that were new to me, I learned about stepping beyond known boundaries and creating new mathematics, I learned about presentation of mathematics, and so on. Not only technical skills and knowledge have developed in this period. As least as important, the fun in mathematics and the challenge and the enchantment one experiences in mathemathical problems still exist. I am grateful to Lex for all this, and I hope we may have a long and fruitful further collaboration.

There are a lot of people to whom I am grateful for their scientific support and personal enthusiasm and encouragements. In particular I would like to mention Bert Gerards, Willem Haemers and also Chris Wildhagen, who, together with Lex, had to listen often to my expositions of unfinished proofs or badly formulated propositions. It is their patience and criticism that brought me on the right track again.

C.A.J. Hurkens

TABLE OF CONTENTS.

Discrete Mathematics 73 (1988/89) 99–109
North-Holland

ON FRACTIONAL MULTICOMMODITY FLOWS AND DISTANCE FUNCTIONS

C.A.J. HURKENS*
Department of Econometrics, Tilburg University, Tilburg, The Netherlands

A. SCHRIJVER
Department of Econometrics, Tilburg University, Tilburg, The Netherlands, and *Mathematical Centre, Kruislaan 413, Amsterdam, The Netherlands*

É. TARDOS**
Department of Computer Science, Eötvös Loránd University, Budapest, Hungary, and *Mathematical Sciences Research, Institute, Berkeley, California, U.S.A.*

Received 13 January 1988

We give some results on the existence of fractional and integral solutions to multicommodity flow problems, and on the related problem of decomposing distance functions into cuts. One of the results is: Let $G = (V, E)$ be a planar bipartite graph. Then there exist subsets $W_1, \ldots, W_t$ of V so that for each pair v', v'' of vertices on the boundary of G, the distance of v' and v'' in G is equal to the number of $j = 1, \ldots, t$ with $|\{v', v''\} \cap W_j| = 1$ and so that the cuts $\delta(W_j)$ are pairwise disjoint.

1. Introduction

In this paper we show some results on fractional and integral multicommodity flows, and on the packing of cuts in planar graphs. Among the results shown is the following:

Let $G = (V, E)$ be a planar bipartite graph. Then there exist subsets $W_1, \ldots, W_t$ of V so that for each pair v', v'' of vertices on the outer face of G, the distance of v' and v'' in G is equal to the number of $j = 1, \ldots, t$ with $|\{v', v''\} \cap W_j| = 1$ and so that the cuts $\delta(W_j)$ are pairwise disjoint

(see Theorem 1 below). Before discussing the results, we first give as a motivation an introduction to multicommodity flows and cut packing, and their 'polarity'.

It is an NP-complete problem to decide if in a given undirected graph $G = (V, E)$, with given pairs of vertices (*ports*) $\{r_1, s_1\}, \ldots, \{r_k, s_k\}$,

there exist k pairwise edge-disjoint paths $P_1, \ldots, P_k$, where P_i connects r_i and s_i $(i = 1, \ldots, k)$ (1)

* Research supported by the Netherlands Organisation for the Advancement of Pure Research (Z.W.O.), through the Stichting Mathematisch Centrum.

** Present address: Dept. of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass. 02139, U.S.A.

(Even et al. [1]). There are however some special cases where good characterizations and polynomial-time algorithms have been found. The larger part of these good characterizations consist of the assertion that the following, obviously necessary, *cut condition* is also sufficient:

$$\text{for each } W \subseteq V : |\delta(W)| \geq |\sigma(W)|\,. \tag{2}$$

Here $\delta(W) := \{e \in E \mid |e \cap W| = 1\}$ and $\sigma(W) := \{i \mid |\{r_i, s_i\} \cap W| = 1\}$. It is easy to see that, if G is connected, we may restrict W in (2) to subsets W for which both W and $V \backslash W$ induce a connected subgraph of G.

Many of these results are restricted to the case where the following *parity condition* holds:

$$\text{for each vertex } v \text{ of } G : |\delta(\{v\})| + |\sigma(\{v\})| \text{ is even.} \tag{3}$$

In one stream of research the given ports are restricted to certain configurations. This stream has begun with the work of Menger [9], Hu [3] and Papernov [12], and has culminated in the work of Lomonosov [7, 8] and Seymour [16]. Lomonosov showed that for any given set of pairs $\{r_1, s_1\}, \ldots, \{r_k, s_k\}$ the following two statements are equivalent:

for each graph $G = (V, E)$ with $V \supseteq \{r_1, s_1, \ldots, r_k, s_k\}$, the cut condition (2) and the parity condition (3) imply (1). (4)

the graph $H := (\{r_1, s_1, \ldots, r_k, s_k\}, \{\{r_1, s_1\}, \ldots, \{r_k, s_k\}\})$ has at most 4 vertices, or is a 5-circuit (possibly with multiple edges), or contains two vertices v', v'' so that $\{r_i, s_i\} \cap \{v', v''\} \neq \emptyset$ for $i = 1, \ldots, k$. (5)

Condition (5) is equivalent to the graph H not having either of the two graphs depicted in Fig. 1 as a subgraph.

Fig. 1.

Lomonosov's theorem implies that if $\{r_1, s_1\}, \ldots, \{r_k, s_k\}$ satisfies (5) and $G = (V, E)$ is a graph with $V \supseteq \{r_1, s_1, \ldots, r_k, s_k\}$, then for any 'capacity' function $c \in \mathbb{Z}_+^E$ and any 'demand' function $d \in \mathbb{Z}_+^k$, the following are equivalent:

there exist paths $P_1^1, \ldots, P_1^{t_1}, P_2^1, \ldots, P_2^{t_2}, \ldots, P_k^{t_k}$ (where each P_i^j connects r_i and s_i, for $i = 1, \ldots, k$, $j = 1, \ldots, t_i$) and rational numbers $\lambda_1^1, \ldots, \lambda_1^{t_1}, \lambda_2^1, \ldots, \lambda_2^{t_2}, \ldots, \lambda_k^{t_k} \geq 0$ so that:

$$\text{(i)} \quad \sum_{j=1}^{t_i} \lambda_j^i = d_i \quad (i=1, \ldots, k),$$

$$\text{(ii)} \quad \sum_{\substack{i=1 \\ e \in P_i^j}}^{k} \sum_{j=1}^{t_i} \lambda_i^j \leq c_e \quad (e \in E). \tag{6}$$

$$\text{for each } W \subseteq V : c(\delta(W)) \geq d(\sigma(W)). \tag{7}$$

(Here $c(F) := \sum_{e \in F} c_e$ for $F \subseteq E$ and $d(J) := \sum_{j \in J} d_j$ for $J \subseteq \{1, \ldots, k\}$.) It is not difficult to see that (6) always implies (7). Conversely, Lomonosov's result implies that if (5) and (7) are satisfied, then we can take each λ_i^j equal to $\frac{1}{2}$ in (6) (by replacing each edge e of G by $2c_e$ parallel edges, and each port $\{r_i, s_i\}$ by $2d_i$ parallel ports).

The assertion:

$$\forall c \in \mathbb{Z}_+^E \; \forall d \in \mathbb{Z}_+^k : (6) \Leftrightarrow (7) \tag{8}$$

is equivalent to the following: Let ε_i denote the ith unit basis vector in $\mathbb{R}^k$, χ^P denote the incidence vector of P in $\mathbb{R}^E$, and ε_e denote the eth unit basis vector in $\mathbb{R}^E$. Then the cone $C \subseteq \mathbb{R}^k \times \mathbb{R}^E$ generated by the vectors:

$$\begin{aligned} &(\varepsilon_i; \chi^P) \quad (i=1, \ldots, k;\ P\ r_i - s_i\text{-path}), \\ &(0; \varepsilon_e) \quad (e \in E) \end{aligned} \tag{9}$$

is determined by the following system of linear inequalities in the vector variable $(d; c) \in \mathbb{R}^k \times \mathbb{R}^E$:

$$\begin{aligned} &d_i \geq 0 && (i=1, \ldots, k), \\ &c_e \geq 0 && (e \in E), \\ &c(\delta(W)) - d(\sigma(W)) \geq 0 && (W \subseteq V). \end{aligned} \tag{10}$$

By polarity (interchanging the roles of generators and constraints), this is equivalent to the assertion that the cone generated by the vectors:

$$\begin{aligned} &(-\chi^{\sigma(W)}; \chi^{\delta(W)}) && (W \subseteq V), \\ &(\varepsilon_i; 0) && (i=1, \ldots, k), \\ &(0; \varepsilon_e) && (e \in E), \end{aligned} \tag{11}$$

(again, for $J \subseteq \{1, \ldots, k\}$, χ^J denotes the incidence vector of J in $\mathbb{R}^k$, while for $J \subseteq E$, χ^J denotes the incidence vector of J in $\mathbb{R}^E$) is determined by the following system of linear inequalities in the vector variable $(m; l) \in \mathbb{R}^k \times \mathbb{R}^E$:

$$\begin{aligned} &m_i + \sum_{e \in P} l_e \geq 0 && (i=1, \ldots, k;\ P\ r_i - s_i\text{-path}), \\ &l_e \geq 0 && (e \in E). \end{aligned} \tag{12}$$

Hence (8) is equivalent to:

for any 'length' function $l: E \to \mathbb{Z}_+$ there exist $W_1, \ldots, W_t \subseteq V$ and

$\mu_1, \ldots, \mu_t \geqslant 0$ so that:

(i) for each $i = 1, \ldots, k$: the minimum length of any $r_i - s_i$-path is at most $\Sigma\,(\mu_j \mid j = 1, \ldots, t;\ i \in \sigma(W_j))$;

(ii) for each $e \in E : l_e \geqslant \Sigma\,(\mu_j \mid j = 1, \ldots, t;\ e \in \delta(W_j))$. (13)

(This can be seen by taking $m_i := -$(minimum length of any $r_i - s_i$-path) in (12).)

Karzanov [4] showed that if (5) holds, then we can take all μ_i equal to $\frac{1}{2}$ in (13). In fact, he showed that (5) is equivalent to:

if $G = (V, E)$ is bipartite and $V \supseteq \{r_1, s_1, \ldots, r_k, s_k\}$, then there exist $W_1, \ldots, W_t \subseteq V$ so that:

(i) for each $i = 1, \ldots, k$: the minimum number of edges in any $r_i - s_i$-path is at most $|\{j = 1, \ldots, t \mid i \in \sigma(W_j)\}|$;

(ii) the cuts $\delta(W_1), \ldots, \delta(W_t)$ are pairwise disjoint. (14)

(13) now follows by replacing each edge e by a path of length $2l_e$. Bipartiteness in (14) is 'dual' to the parity condition (3).

A second stream of research restricts G to planar graphs. First, Okamura and Seymour [11] showed that the cut condition (2) and the parity condition (3) imply (1) if:

G is planar, and all $r_1, s_1, \ldots, r_k, s_k$ are vertices on the boundary of G. (15)

Okamura [10] extended this result by relaxing (15) to:

G is planar, and there exist faces I and O (where we can assume O to be the outer face, without loss of generality), so that for each $i = 1, \ldots, k : r_i, s_i \in I$ or $r_i, s_i \in O$. (16)

Seymour [17] showed that (2) and (3) imply (1) if:

the graph $(V, E \cup \{\{r_1, s_1\}, \ldots, \{r_k, s_k\}\})$ is planar. (17)

In Oberwolfach the following extension of the Okamura-Seymour theorem, due to Van Hoesel and Schrijver [2], conjectured by Kurt Mehlhorn, was presented:

Let $G = (V, E)$ be a planar graph. Let O and I be the outer and some other fixed face. Let $C_1, \ldots, C_k$ be curves in $\mathbb{R}^2 \backslash (I \cup O)$, with end points being vertices on $I \cup O$, so that for each vertex v of G the degree of v in G has the same parity as the number of curves C_i beginning or ending in v (counting a curve beginning and ending in v for two). Then there exist pairwise edge-disjoint paths $P_1, \ldots, P_k$ in G so that P_i is homotopic to C_i in the space $\mathbb{R}^2 \backslash (I \cup O)$ for $i = 1, \ldots, k$, if and only if for each path Q in the dual graph of G from I or O to I or O, the number of edges in Q is not smaller than the number of times Q necessarily intersects the curves C_i. (18)

With this last number we mean $\sum_{i=1}^{k}(\min\{|D \cap Q| \mid D$ homotopic to $C_i\})$. Mehlhorn's conjecture was motivated by work on grid graphs (cf. [6]), related to the problem of the automatic design of integrated circuits. It is not difficult to see that (18) implies the Okamura-Seymour theorem.

In this contribution to the Proceedings, we discuss some problems, observations and results related to the above, which were inspired by discussions we had in Oberwolfach.

2. Distance functions in planar graphs

In the same manner as (13) (under condition (5)) follows from Lomonosov's theorem, by considering cones one can derive the following from the Okamura-Seymour theorem: Let $G=(V, E)$ be a planar graph, and let $l: E \to \mathbb{Z}_+$ be a 'length' function. Then there exist subsets $W_1, \ldots, W_t$ of V and $\mu_1, \ldots, \mu_t \geq 0$ so that:

(i) for each pair v', v'' of vertices on the boundary of G the minimum length of any v'-v''-path is at most $\sum(\mu_j \mid j=1, \ldots, t;\ |\{v', v''\} \cap W_j|=1)$;

(ii) for each $e \in E: l(e) \geq \sum(\mu_j \mid j=1, \ldots, t;\ e \in \delta(W_j))$. (19)

In fact, we can take each μ_j equal to $\frac{1}{2}$, as follows from the following theorem:

Theorem 1. *Let $G=(V, E)$ be a planar bipartite graph. Then there exist subsets $W_1, \ldots, W_t$ of V so that for each pair v', v'' of vertices on the boundary of G, the minimum number of edges in any v'-v''-path is equal to the number of $j=1, \ldots, t$ with $|\{v', v''\} \cap W_j|=1$ and so that the cuts $\delta(W_j)$ are pairwise disjoint.*

We show how this theorem can be derived from the Okamura-Seymour theorem. First, let $C=(V, E)$ be a circuit with k vertices and k edges:

$$\begin{aligned} V &= \{v_1, \ldots, v_k\}, \\ E &= \{e_1=\{v_0, v_1\}, \ldots, e_k=\{v_{k-1}, v_k\}\}, \end{aligned} \tag{20}$$

where $v_0=v_k$. Let $\binom{V}{2}$ and $\binom{E}{2}$ denote the set of undirected pairs of elements from V and E, respectively. Let M be the $\binom{V}{2} \times \binom{E}{2}$ matrix given by:

$$\begin{aligned} M_{\{v_i, v_j\}, \{e_g, e_h\}} &= 1 \quad \text{if } \{v_i, v_j\} \text{ and } \{e_g, e_h\} \text{ "cross"}; \\ &= 0 \quad \text{otherwise}, \end{aligned} \tag{21}$$

where $\{v_i, v_j\}$ and $\{e_g, e_h\}$ are said to cross if v_i and v_j belong to different components of the graph $C \backslash \{e_g, e_h\}$. We show that the matrix M is nonsingular, with $\binom{E}{2} \times \binom{V}{2}$ inverse N given by:

$$\begin{aligned} N_{\{e_g, e_h\}, \{v_i, v_j\}} &= +\tfrac{1}{2} \quad \text{if } \{v_i, v_j\}=\{v_g, v_h\} \quad \text{or} \quad \{v_i, v_j\}=\{v_{g-1}, v_{h-1}\}, \\ &= -\tfrac{1}{2} \quad \text{if } \{v_i, v_j\}=\{v_g, v_{h-1}\} \quad \text{or} \quad \{v_i, v_j\}=\{v_{g-1}, v_h\}, \\ &= 0 \qquad \text{otherwise}. \end{aligned} \tag{22}$$

Proposition. $N = M^{-1}$.

Proof. Choose $\{e_g, e_h\}, \{e_a, e_b\} \in \binom{E}{2}$. Then

$$(NM)_{\{e_g,e_h\},\{e_a,e_b\}} = \tfrac{1}{2}M_{\{v_g,v_h\},\{e_a,e_b\}} + \tfrac{1}{2}M_{\{v_{g-1},v_{h-1}\},\{e_a,e_b\}} - \tfrac{1}{2}M_{\{v_g,v_{h-1}\},\{e_a,e_b\}} - \tfrac{1}{2}M_{\{v_{g-1},v_h\},\{e_a,e_b\}}. \tag{23}$$

If $\{g, h\} = \{a, b\}$ then it is easy to see that this last expression is equal to 1. If $\{g, h\} \neq \{a, b\}$, then without loss of generality $g \notin \{a, b\}$. Then

$$M_{\{v_g,v_h\},\{e_a,e_b\}} = M_{\{v_{g-1},v_h\},\{e_a,e_b\}} \quad \text{and}$$
$$M_{\{v_g,v_{h-1}\},\{e_a,e_b\}} = M_{\{v_{g-1},v_{h-1}\},\{e_a,e_b\}}, \tag{24}$$

which implies that (23) is 0. □

[It can be shown that $|\det M| = 2^{\binom{k-1}{2}}$.]

Proof of Theorem 1. Without loss of generality, G is 2-connected. Let $v_1, \ldots, v_k$ be the vertices on the boundary of G in order, and let $e_1 = \{v_0, v_1\}, \ldots, e_k = \{v_{k-1}, v_k\}$ be the edges on the boundary of G (where $v_0 := v_k$). Let M and N be the matrices as above with respect to the circuit $(W := \{v_1, \ldots, v_k\},\ F := \{e_1, \ldots, e_k\})$. Let $m : \binom{W}{2} \to \mathbb{Z}_+$ be defined by: $m(\{v_i, v_j\}) :=$ minimum number of edges in any $v_i - v_j$-path. Let $d := Nm$. Since G is planar and bipartite, Nm is a nonnegative integer vector. In fact, for each $g = 1, \ldots, k$:

$$\sum_{\substack{h=1\\ h\neq g}}^{k} d_{\{e_g,e_h\}} = m_{\{v_{g-1},v_g\}} = 1, \tag{25}$$

as easily follows from the definition of N (or from $Md = m$). Therefore, for each $g \in \{1, \ldots, k\}$ there is a unique $h \neq g$ such that $d_{\{e_g,e_h\}} = 1$, i.e. the collection $\{\{e_g, e_h\} \mid d_{\{e_g,e_h\}} = 1\}$ partitions $\{e_1, \ldots, e_k\}$.

Now let G^* be the (planar) dual graph of G. Put a new vertex w_g on every edge e_g^* of G^* corresponding to edge e_g of G, and next delete the vertex of G^* corresponding to the unbounded face, together with all edges incident with it. Call the graph thus obtained H.

So the collection $\{\{w_g, w_h\} \mid d_{\{e_g,e_h\}} = 1\}$ partitions $\{w_1, \ldots, w_k\}$. Let these pairs be the ports for H. Since each w_g has degree 1 in H, the parity condition (3) is satisfied. Also the cut condition (2) is satisfied. Indeed, let Z be a subset of the vertex set Y of H so that both Z and $Y \backslash Z$ induce a connected subgraph of H. We may assume that there exist g and h so that $w_{g+1}, w_h \in Z$ and $w_g, w_{h+1} \notin Z$. Then

$$|\delta(Z)| \geq m_{\{v_g,v_h\}} = (Md)_{\{v_g,v_h\}} = |\sigma(Z)|. \tag{26}$$

So the cut condition is satisfied.

Hence, by the Okamura-Seymour theorem, there exist pairwise edge-disjoint paths $Q_1, \ldots, Q_{\frac{1}{2}k}$ in H connecting the ports. In G this gives pairwise edge-disjoint cuts $\sigma(W_1), \ldots, \sigma(W_{\frac{1}{2}k})$ so that for any g, h, if $d_{\{e_g, e_h\}} = 1$, then $e_g, e_h \in \delta(W_j)$ for some j. Hence for all i, j:

$$\begin{aligned} m_{\{v_i, v_j\}} = (Md)_{\{v_i, v_j\}} &= \sum_{\{e_g, e_h\} \in \binom{E}{2}} M_{\{v_i, v_j\}, \{e_g, e_h\}} d_{\{e_g, e_h\}} \\ &= |\{f = 1, \ldots, \tfrac{1}{2}k \mid |\{v_i, v_j\} \cap W_f| = 1\}| . \qquad (27) \quad \square \end{aligned}$$

The above reasoning also implies that for any planar bipartite graph G there is a unique partitioning of the edges on the boundary C into pairs $\pi_1, \ldots, \pi_{\frac{1}{2}k}$ of edges so that for any two vertices v', v'' on the boundary of G, the distance from v' to v'' in G is equal to the number of pairs π_j which cross (i.e. separate) v' and v'' on C.

Another application of the above proposition is the following. Let $C = (V, E)$ be a circuit (satisfying (20)). Call a function $m : \binom{V}{2} \to \mathbb{R}_+$ *realizable as a distance function of a planar graph with boundary C*, or briefly *realizable*, if there exists a planar graph $G = (V', E')$, with $V' \supseteq V$, $E' \supseteq E$ and with boundary C, and a length function $l : E \to \mathbb{R}_+$ so that for all $v', v'' \in V$, $m(\{v', v''\})$ is equal to the minimum length of any $v' - v''$-path in G.

Theorem 2. *A function $m : \binom{V}{2} \to \mathbb{R}_+$ is realizable, if and only if for all $i, j = 1, \ldots, k$ we have $m(\{v_i, v_j\}) + m(\{v_{i-1}, v_{j-1}\}) \geq m(\{v_i, v_{j-1}\}) + m(\{v_{i-1}, v_j\})$ (taking $m(\{v_i\}) := m(\{v_j\}) := 0$).*

Proof. Necessity being trivial, we show sufficiency. We construct a graph G as follows. Let $w_1, \ldots, w_k = w_0$ be points on the unit circle (in the cyclic order given). Add all line-segments $\overline{w_g w_h}$ ($g, h = 1, \ldots, k$; $g \neq h$). Let W be the set of points which are on two or more of these line-segments. Clearly, the figure now forms a planar graph H, with vertex set W. Let H^* be the dual graph. Put a new point v_i on the edge of H^* corresponding to edge $\overline{w_i w_{i+1}}$ of H ($i = 0, \ldots, k-1$), delete the vertex of H^* corresponding to the outer face of H, and delete all edges incident to it. Moreover, add edges $e_1 = \{v_0, v_1\}, \ldots, e_k = \{v_{k-1}, v_k\}$ (where $v_k := v_0$). This makes the graph $G = (V', E')$.

The condition in the theorem states that $d := Nm \geq 0$. For each edge e of G define $l(e) := d(\{e_g, e_h\})$ if e corresponds to an edge in H which is on the line-segment $\overline{w_g w_h}$, while $l(e) := \infty$ (or big enough, or $m(\{v_{i-1}, v_i\})$) if $e = e_i = \{v_{i-1}, v_i\}$ for some i.

It is easy to see (using the fact that $Md = m$) that this gives a realization as required. $\square$

3. Two counterexamples

In Okamura's theorem (cf. (16)) we generally cannot accept 'mixed' ports, i.e. ports $\{r_i, s_i\}$ with $r_i \in O$ and $s_i \in I$, as is shown by the following example of

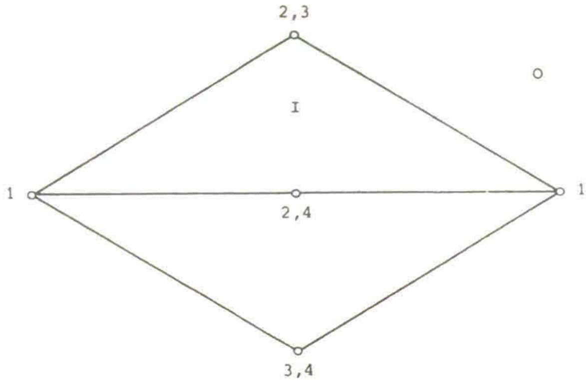


Fig. 2.

Okamura (Fig. 2). In this example (denoting r_i and s_i just by i), the cut condition (2) and the parity condition (3) are satisfied, but there are no paths as required, since each $r_i - s_i$-path has at least two edges, while there are six edges in total.

This last argument shows that there does not even exist a 'fractional' solution, in the sense of (6) (taking $c \equiv 1$, $d \equiv 1$). András Frank asked whether the

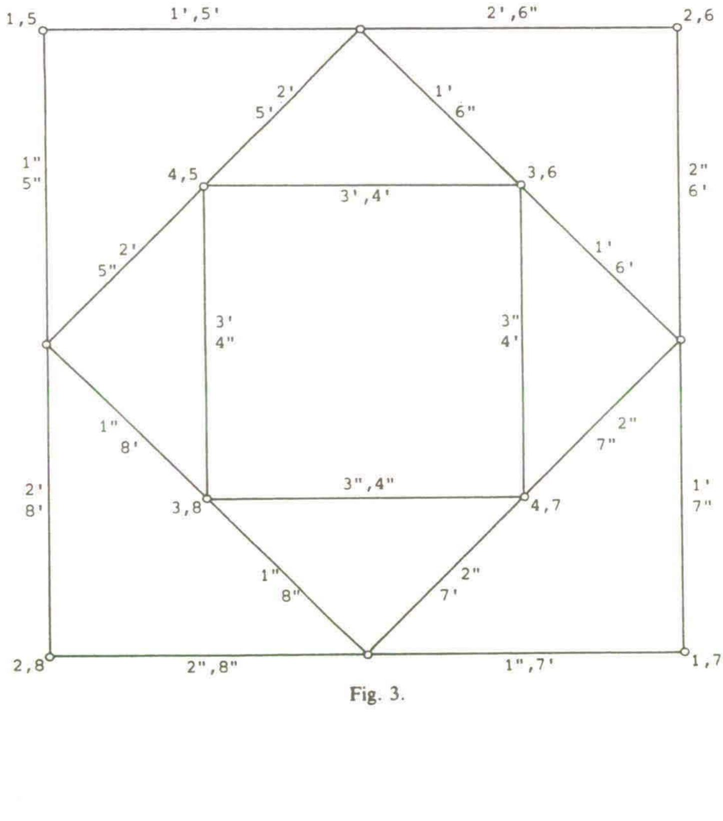


Fig. 3.

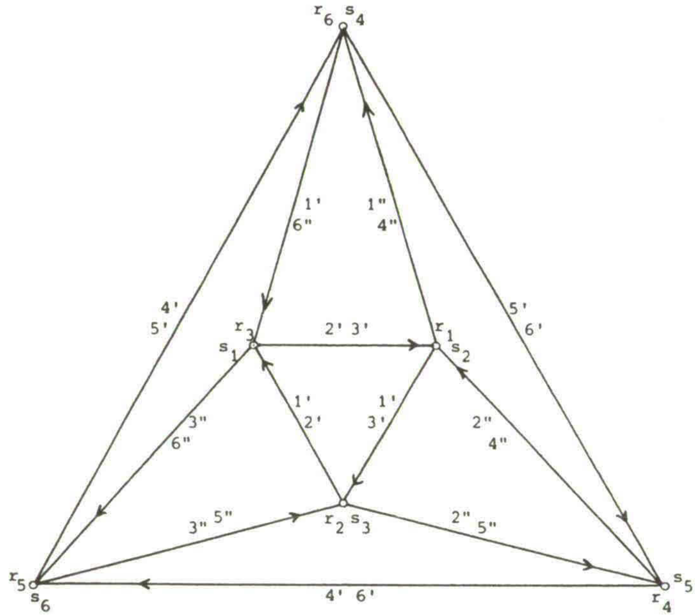


Fig. 4.

existence of such a fractional solution might imply the existence of paths as required. A negative answer is provided by the example in Fig. 3. Note that the parity condition is satisfied. For each $i = 1, \ldots, 8$, the two paths indicated by i' and i'' are $i - i$-paths. Each edge is in exactly two of these paths. So this yields a fractional solution in the sense of (7) (with all λ_i^j equal to $\frac{1}{2}$). However, there is no integer solution, i.e. (1) is not fulfilled. For suppose $P_1, \ldots, P_8$ are pairwise edge-disjoint paths, with P_i connecting r_i and s_i $(i = 1, \ldots, 8)$. Clearly, $|P_i| \geqslant 4$ for $i = 1, 2$, and $|P_i| \geqslant 2$ for $i = 3, \ldots, 8$. Moreover, $|P_1| + \cdots + |P_8| \leqslant 20$, since there are 20 edges. Hence $|P_3| = |P_4| = 2$. But there do not exist edge-disjoint $r_3 - s_3$- and $r_4 - s_4$-paths, both of length 2.

The second example also answers a question of András Frank, concerning a directed analogue of Seymour's theorem (cf. (17)). Consider the directed graph shown in Fig. 4. It is easy to see that there are no pairwise arc-disjoint directed paths $P_1, \ldots, P_6$ so that P_i is an $r_i - s_i$-path $(i = 1, \ldots, 6)$. Note that in each vertex v, indegree $(v) + |\{i \mid s_i = v\}| =$ outdegree $(v) + |\{i \mid r_i = v\}|$ (the analogue of the parity condition). There exists a 'fractional' solution: for $i = 1, \ldots, 6$, the paths indicated by i' and i'' form two $r_i - s_i$-paths, while each arc is in exactly two of these paths (it follows that the directed analogue of the cut condition is satisfied).

4. Some further notes

We mention some questions. Is there a common generalization of the Okamura and the Van Hoesel–Schrijver theorem (cf. (16) and (18))? Or can one be derived from the other? Note that in order to derive the Okamura theorem from (18) it suffices to show that, given the input of the Okamura theorem, one can specify curves connecting r_i and s_i $(i = 1, \ldots, k)$ in $\mathbb{R}^2 \backslash (I \cup O)$ so that the condition mentioned in (18) is satisfied. We do not see a direct way (i.e. one not using the Okamura theorem itself) to derive this.

In [13] Theorem 1 is extended to the case where we also allow that both v' and v'' belong to some other fixed face I. This corresponds to the Okamura theorem, in the same way as Theorem 1 corresponds to the Okamura–Seymour theorem. Karzanov [5] observed that a similar result with respect to Seymour's theorem (cf. (17)) can be derived from Seymour's results on 'sums of circuits' [15].

The Van Hoesel–Schrijver theorem (18) cannot be extended in the obvious way to the case where there are more 'holes', as is shown by the example in Fig. 5.

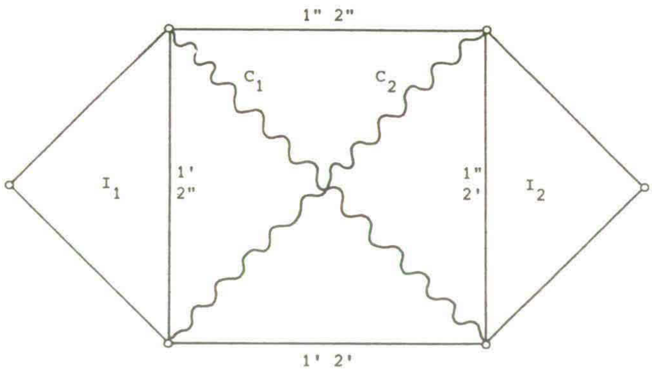


Fig. 5.

Here the "dual curve condition" given in (18) is satisfied, but there are no edge-disjoint paths P_1 and P_2, where P_i is homotopic to C_i in the space $\mathbb{R}^2$ $(O \cup I_1 \cup I_2)$. However, there is a 'fractional' solution, by taking each of the paths $1', 1'', 2', 2''$ with multiplicity $\frac{1}{2}$. In Oberwolfach, Professor Crispin Nash-Williams asked whether the dual curve condition implies the existence of a fractional solution (in any planar graph with any number of holes). This question can be answered affirmatively, as will be shown in a forthcoming paper [14].

Acknowledgements

We thank András Frank, Bert Gerards and an anonymous referee for very helpful discussions and suggestions.

References

[1] S. Even, A. Itai and A. Shamir, On the complexity of time-table and multicommodity flow problems, SIAM Journal on Computing 5 (1976) 691–703.

[2] C. van Hoesel and A. Schrijver, Edge-disjoint homotopic paths in a planar graph with one hole, Report OS-R8608 (Centrum voor Wiskunde en Informatica, Amsterdam, 1986), to appear in J. Combinat. Theory (B).

[3] T.C. Hu, Multi-commodity network flows, Operations Research 11 (1963) 344–360.

[4] A.V. Karzanov, Metrics and undirected cuts, Mathematical Programming 32 (1985) 183–198.

[5] A.V. Karzanov, personal communication (1987).

[6] M. Kaufmann and K. Mehlhorn, On local routings of two-terminal nets, preprint 03/1986 (Fachbereich 10 Informatik, Universität des Saarlandes, Saarbrücken, 1986).

[7] M.V. Lomonosov, Solutions for two problems on flows in networks, submitted to Problemy Peredači Informacii (1976).

[8] M.V. Lomonosov, Multiflow feasibility depending on cuts, Graph Theory Newsletter 9 (1) (1979) 4.

[9] K. Menger, Zur allgemeinen Kurventheorie, Fundamenta Mathematicae 10 (1927) 96–115.

[10] H. Okamura, Multicommodity flows in graphs, Discrete Applied Mathematics 6 (1983) 55–62.

[11] H. Okamura and P.D. Seymour, Multicommodity flows in planar graphs, Journal of Combinatorial Theory (B) 31 (1981) 75–81.

[12] B.A. Papernov, On existence of multicommodity flows (in Russian), in: Studies in Discrete Optimization (A.A. Fridman, ed.) (Nauka, Moscow, 1976) 230–261.

[13] A. Schrijver, Distances and cuts in planar graphs, Report OS-R8610 (Centrum voor Wiskunde en Informatica, Amsterdam, 1986), to appear in J. Combinat. Theory (B).

[14] A. Schrijver, Decomposition of graphs on surfaces and a homotopic circulation theorem, Report OS-R8719 (Centrum voor Wiskunde en Informatica, Amsterdam 1987).

[15] P.D. Seymour, Sums of circuits, in: Graph Theory and Related Topics (J.A. Bondy and U.S.R. Murty, eds.) (Academic Press, New York, 1978) 341–355.

[16] P.D. Seymour, Four-terminus flows, Networks 10 (1980) 79–86.

[17] P.D. Seymour, On odd cuts and plane multicommodity flows, Proceedings of the London Mathematical Society (3) 42 (1981) 178–192.

SIAM J. DISC. MATH.
Vol. 2, No. 1, pp. 68–72, February 1989

ON THE SIZE OF SYSTEMS OF SETS EVERY t OF WHICH HAVE AN SDR, WITH AN APPLICATION TO THE WORST-CASE RATIO OF HEURISTICS FOR PACKING PROBLEMS*

C. A. J. HURKENS† AND A. SCHRIJVER‡

Abstract. Let $E_1, \cdots, E_m$ be subsets of a set V of size n, such that each element of V is in at most k of the E_i and such that each collection of t sets from $E_1, \cdots, E_m$ has a system of distinct representatives (SDR). It is shown that $m/n \leqq (k(k-1)^r - k)/(2(k-1)^r - k)$ if $t = 2r - 1$, and $m/n \leqq (k(k-1)^r - 2)/(2(k-1)^r - 2)$ if $t = 2r$. Moreover it is shown that these upper bounds are the best possible. From these results the "worst-case ratio" of certain heuristics for the problem of finding a maximum collection of pairwise disjoint sets among a given collection of sets of size k is derived.

Key words. packing, system of distinct representatives, worst-case ratio, heuristics

AMS(MOS) subject classifications. 05C65, 05A05, 90C27

1. Introduction. We prove the following theorem, where m, n, k, and t are positive integers, with $k \geqq 3$.

THEOREM 1. *Let $E_1, \cdots, E_m$ be subsets of the set V of size n, such that we have the following*:

(1) (i) *Each element of V is contained in at most k of the sets $E_1, \cdots, E_m$;*
(ii) *Any collection of at most t sets among $E_1, \cdots, E_m$ has a system of distinct representatives.*

Then, we have the following:

$$(2)\quad \text{(i)}\ \frac{m}{n} \leqq \frac{k(k-1)^r - k}{2(k-1)^r - k} \quad \text{if } t = 2r-1;$$

$$\text{(ii)}\ \frac{m}{n} \leqq \frac{k(k-1)^r - 2}{2(k-1)^r - 2} \quad \text{if } t = 2r.$$

Note that by the König–Hall Theorem, condition (1)(ii) can be replaced by the following:

(3) For any $s \leqq t$, any s of the sets among $E_1, \cdots, E_m$ cover at least s elements of V.

We give a proof of Theorem 1 in § 2. We also show that the bounds given in (2) are best possible in the following sense.

THEOREM 2. *For any fixed k, t (with $k \geqq 3$), there exist m, n and $E_1, \cdots, E_m \subseteq V$ (with $|V| = n$) satisfying* (1) *and having equality in the appropriate line of* (2).

The proof of Theorem 2 is based on a construction using regular graphs of large girth (see § 3).

Finally, in § 4 we apply these results to derive the worst-case ratio of certain heuristic algorithms for the problem of finding a largest family of pairwise disjoint sets among a given family of sets of size k (this problem is NP-complete for any $k \geqq 3$).

---

* Received by the editors April 1, 1988; accepted for publication July 13, 1988. Part of this research was performed while the authors were visiting the Rutgers Center for Operations Research, Rutgers University, New Brunswick, New Jersey 08903.

† Department of Econometrics, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, the Netherlands. The research of this author was supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) through the Stichting Mathematisch Centrum.

‡ Department of Econometrics, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, the Netherlands, and Mathematical Centre, Kruislaan 413, 1098 SJ Amsterdam, the Netherlands.

2. Proof of Theorem 1. To show Theorem 1, we first give a lemma. Let $E_1, \cdots, E_m$ be a collection of finite nonempty sets, which we order so that $|E_1|, \cdots, |E_h| \geqq 2$ and $|E_{h+1}| = \cdots = |E_m| = 1$, for some $h \leqq m$. We define a new collection as follows. Let

$$W := E_{h+1} \cup \cdots \cup E_m. \tag{4}$$

Let for each $i = 1, \cdots, h$, X_i be a set of size $|E_i| - 2$, disjoint from $E_1 \cup \cdots \cup E_m$ and so that if $i \neq j$ then $X_i \cap X_j = \varnothing$. Let $X_1 \cup \cdots \cup X_h =: \{y_1, \cdots, y_q\}$. Then the *derived* collection of sets is formed by the following sets:

$$(E_1 \backslash W) \cup X_1, \cdots, (E_h \backslash W) \cup X_h, \{y_1\}, \cdots, \{y_q\}. \tag{5}$$

Furthermore, we define a collection $E_1, \cdots, E_m$ to have the *t-SDR-property* if any t sets among $E_1, \cdots, E_m$ have a system of distinct representatives.

LEMMA. *For $t \geqq 3$, if $E_1, \cdots, E_m$ has the t-SDR-property, then the derived collection (5) has the $(t-2)$-SDR-property.*

Proof. Suppose (5) does not have the $(t-2)$-SDR-property. Then there exists a collection Π of p sets among (5) covering at most $p-1$ elements, for some $p \leqq t-2$. Assume we have chosen p minimal. This immediately implies the following:

(6) (i) $|\cup \Pi| = p - 1$;
(ii) Each element in $\cup \Pi$ is covered by at least two sets in Π.

From (6)(ii) we directly have for any $i = 1, \cdots, h$ and $x \in X_i$:

$$\{x\} \in \Pi \Leftrightarrow (E_i \backslash W) \cup X_i \in \Pi. \tag{7}$$

Without loss of generality, all sets $(E_1 \backslash W) \cup X_1, \cdots, (E_h \backslash W) \cup X_h$ belong to Π (as we can delete all sets E_j from $E_1, \cdots, E_h$ for which $(E_j \backslash W) \cup X_j \notin \Pi$), and without loss of generality, $(E_1 \cup \cdots \cup E_h) \cap W = E_{h+1} \cup \cdots \cup E_m$.

Note the following:

$$q = |X_1 \cup \cdots \cup X_h| = \sum_{i=1}^{h} (|E_i| - 2), \qquad p = h + q, \tag{8}$$

$$\left| \bigcup_{i=1}^{h} (E_i \backslash W) \right| = |\cup \Pi| - q = (p-1) - q = h - 1.$$

So,

$$\left| \bigcup_{i=1}^{m} E_i \right| = \left| \bigcup_{i=1}^{h} (E_i \cap W) \right| + \left| \bigcup_{i=1}^{h} (E_i \backslash W) \right| = (m-h) + (h-1) = m - 1. \tag{9}$$

Moreover, by (6)(ii), $\sum_{i=1}^{h} |E_i \backslash W| \geqq 2 \cdot |\bigcup_{i=1}^{h} (E_i \backslash W)|$, and hence

$$\begin{aligned} m &= h + \left| \bigcup_{i=1}^{h} (E_i \cap W) \right| \leqq h + \sum_{i=1}^{h} |E_i \cap W| = h + \sum_{i=1}^{h} |E_i| - \sum_{i=1}^{h} |E_i \backslash W| \\ &\leqq h + \sum_{i=1}^{h} |E_i| - 2 \cdot \left| \bigcup_{i=1}^{h} (E_i \backslash W) \right| = h + 2h + \sum_{i=1}^{h} (|E_i| - 2) - 2(h-1) \\ &= h + 2h + q - 2(h-1) = h + q + 2 = p + 2 \leqq t. \end{aligned} \tag{10}$$

Inequalities (9) and (10) contradict the fact that $E_1, \cdots, E_m$ has the t-SDR-property. □

Proof of Theorem 1. We prove Theorem 1 by induction on t.

Case 1. $t = 1$. Then we have that each of $E_1, \cdots, E_m$ is nonempty, and hence $m \leqq \sum_{i=1}^{m} |E_i| \leqq kn$, by (1)(i).

Case 2. $t = 2$. Then we have that each of $E_1, \cdots, E_m$ is nonempty, and that no two of the singletons among $E_1, \cdots, E_m$ are the same. Without loss of generality, let $E_{h+1}, \cdots, E_m$ be the singletons among $E_1, \cdots, E_m$. Then $m - h \leqq n$, and

$$m+h=2h+(m-h)\leqq \sum_{i=1}^{h} |E_i|+\sum_{i=h+1}^{m} |E_i| = \sum_{i=1}^{m} |E_i| \leqq kn \tag{11}$$

(by (1)(i)). Hence $2m = (m - h) + (m + h) \leqq (k + 1)n$, and (2) follows.

Case 3. $t \geqq 3$. Then consider the derived collection $E'_1, \cdots, E'_{m'}$ on $V' := \bigcup_{i=1}^{m'} E'_i$ as in (5). Note that $m' = h + q$ and $n' := |V'| = n - |W| + q$. Denote the right-hand side term in (2) by $\varphi(k, t)$.

As by the lemma above, $E'_1, \cdots, E'_{m'}$ has the $(t - 2)$-SDR-property, and as trivially each element of V' is in at most k of the sets $E'_1, \cdots, E'_{m'}$ we have by induction that $m' \leqq \varphi(k, t - 2)n'$. That is,

$$h+q\leqq\varphi(k,t-2)(n-|W|+q). \tag{12}$$

Writing the terms in different order, we have

$$\varphi(k,t-2)|W|+h-(\varphi(k,t-2)-1)q\leqq\varphi(k,t-2)n. \tag{13}$$

Moreover, as $E_1, \cdots, E_m$ cover any element at most k times:

$$|W|+2h+q=|W|+2h+\sum_{i=1}^{h}(|E_i|-2)=|W|+\sum_{i=1}^{h}|E_i|=\sum_{i=1}^{m}|E_i|\leqq kn. \tag{14}$$

Hence,

$$\begin{aligned} m&=h+|W|\\ &=\frac{1}{2\varphi(k,t-2)-1}(\varphi(k,t-2)|W|+h-(\varphi(k,t-2)-1)q)\\ &\quad+\frac{\varphi(k,t-2)-1}{2\varphi(k,t-2)-1}(|W|+2h+q)\\ &\leqq\frac{1}{2\varphi(k,t-2)-1}\varphi(k,t-2)n+\frac{\varphi(k,t-2)-1}{2\varphi(k,t-2)-1}kn\\ &=\frac{(k+1)\varphi(k,t-2)-k}{2\varphi(k,t-2)-1}n=\varphi(k,t)n. \end{aligned} \tag{15}$$

The last equality follows directly by substituting the corresponding right-hand side of (2). □

3. Proof of Theorem 2. To prove Theorem 2 we use a result of Erdös and Sachs [1]:

(16) For every k and γ there exists a k-regular graph of girth γ.

As a consequence of (16) we have the following:

(17) For every k, s, and γ there exists a bipartite graph of girth at least γ, with color classes U and W, say, such that each vertex in U has degree k, and each vertex in W has degree s.

(To see that (17) follows from (16), let H be a $2ks$-regular graph of girth γ. Consider any Eulerian orientation of the edges of H (i.e., one for which all indegrees and outdegrees equal ks). Split each vertex v into $k + s$ vertices $v_1, \cdots, v_k, w_1, \cdots, w_s$ and divide the arcs entering v equally over $v_1, \cdots, v_k$ and divide the arcs leaving v equally over $w_1, \cdots, w_s$. Forgetting the orientations, we obtain a bipartite graph with the required properties.)

Now choose k, t. Let $r := \lfloor \frac{1}{2}t \rfloor$. Consider the tree T, with vertices $1, 2, \cdots, 1 + (k-1) + (k-1)^2 + \cdots + (k-1)^{r-1}$, so that for $i < j$, vertices i and j are connected by an edge, if and only if $(k-1)i \leqq j \leqq (k-1)i + (k-2)$. So each vertex has degree k, except for vertex 1, which has degree $k - 1$, and for the vertices $1 + (k-1) + \cdots + (k-1)^{r-2} + 1, \cdots, 1 + (k-1) + \cdots + (k-1)^{r-1}$, which have degree one.

First let t be even. Let G be a $(k-1)^r$-regular graph of girth $t + 1$ (cf. (16)). Let G have p vertices: $v_1, \cdots, v_p$. Consider p copies $T_1, \cdots, T_p$ of T (denoting the copy of vertex i in T_j by i_j). For each $j = 1, \cdots, p$, partition the set of $(k-1)^r$ edges of G incident to v_j (arbitrarily) into $(k-1)^{r-1}$ classes of size $k - 1$, and connect them to the $(k-1)^{r-1}$ vertices i_j in T_j of degree one. So the final graph $H = (W, F)$ has all degrees equal to k, except for the vertices $1_1, \cdots, 1_p$, which have degree $k - 1$. Let $E_1, \cdots, E_m$ be the collection $F \cup \{\{1_1\}, \cdots, \{1_p\}\}$. This collection clearly satisfies (1)(i), and direct counting shows equality in (2)(ii). To see that the collection satisfies (1)(ii), let $E_1, \cdots, E_s$ form a subcollection with $|E_1 \cup \cdots \cup E_s| < s$ and s as small as possible. Suppose $s \leqq t$. As $E_1, \cdots, E_s$ must form a connected hypergraph, it contains at most one singleton (since any path between 1_i and 1_j in H contains at least $t - 1$ edges). So assume $E_2, \cdots, E_s$ are edges of H. Then they do not contain any circuit (as each T_i is a tree and as G has girth $t + 1 > s$). So $|E_2 \cup \cdots \cup E_s| \geqq s$, a contradiction.

Next let t be odd. Let G be a bipartite graph, of girth at least $t + 1$, so that in one color class U each vertex has degree $(k-1)^r$ and in the other color class W each vertex has degree k. Let $U =: \{u_1, \cdots, u_p\}$. Consider again p copies $T_1, \cdots, T_p$ of T, as above. For $j = 1, \cdots, p$ partition the set of $(k-1)^r$ edges of G incident to u_j (arbitrarily) into $(k-1)^{r-1}$ classes of size $k - 1$, and connect them to the $(k-1)^{r-1}$ vertices i_j in T_j of degree one. Again, the final graph $H = (W, F)$ has all degrees equal to k, except for the vertices $1_1, \cdots, 1_p$ that have degree $k - 1$. Let $E_1, \cdots, E_m$ be the collection $F \cup \{\{1_1\}, \cdots, \{1_p\}\}$. Similarly, as above, we show that this collection satisfies (1) and has equality in (2)(i).

4. Application to the worst-case ratio of heuristics. The problem of finding a largest collection of pairwise disjoint sets among a given collection $X_1, \cdots, X_q$ of k-sets is NP-complete, for any $k \geqq 3$. Call any collection of pairwise disjoint sets a *packing*.

For any fixed s, we can apply the following heuristic algorithm H_s. Start with the empty packing. If we have found a packing $Y_1, \cdots, Y_n$ from $X_1, \cdots, X_q$, we could select $p \leqq s$ sets among $Y_1, \cdots, Y_n$, and replace them by $p + 1$ sets from $X_1, \cdots, X_q$, so that the arising collection is a packing with $n + 1$ sets. Repeating this, the algorithm terminates with a collection $Y_1, \cdots, Y_n$ so that

(18) For each $p \leqq s$, the union of any $p + 1$ pairwise disjoint sets among $X_1, \cdots, X_q$ intersects at least $p + 1$ sets among $Y_1, \cdots, Y_n$.

This defines heuristic H_s, which is, for any fixed s, a polynomial-time algorithm—however it clearly need not lead to a largest packing. We might ask how far the packing found with H_s is from the largest packing.

To this end, consider a largest packing $Z_1, \cdots, Z_m$ from $X_1, \cdots, X_q$. We claim that m/n satisfies the bounds given in (2), taking $t := s + 1$, and that these bounds are best possible. That is, the "worst-case ratio" of the heuristic is given in (2).

Indeed, let

$$(19)\qquad V:=\{Y_1,\cdots,Y_n\}\quad\text{and}\quad E_i:=\{Y_j\mid Y_j\cap Z_i\neq 0\}\quad\text{for } i=1,\cdots,m.$$

Then by (18), $E_1, \cdots, E_m$ satisfy (1), and hence we obtain the bounds given in (2).

In turn, it is not difficult to see that for any collection $E_1, \cdots, E_m$ of sets of size at most k, containing any point at most k times, we can assume they are of form (19) for certain packings $Y_1, \cdots, Y_n$ and $Z_1, \cdots, Z_m$ of k-sets. Thus starting with $E_1, \cdots, E_m$ as described in § 3 above, making these $Y_1, \cdots, Y_n, Z_1, \cdots, Z_m$, and taking $\{X_1, \cdots, X_q\} := \{Y_1, \cdots, Y_n, Z_1, \cdots, Z_m\}$, we obtain a system of sets attaining the worst-case ratio. (That is because we may assume that H_s selects the sets $Y_1, \cdots, Y_n$ in the first n iterations.)

Note that we may assume even that the sets $Y_1, \cdots, Y_n, Z_1, \cdots, Z_m$ form the collection of all cliques of size k in a graph. Hence, we cannot obtain a better worst-case ratio by restricting the collections of sets to collections of k-cliques.

REFERENCE

[1] P. Erdös and H. Sachs, *Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl*, Wiss. Z. Univ. Halle, Math.-Nat., 12 (1963), pp. 251–258.

ON THE DIAMETER OF THE EDGE COVER POLYTOPE.

C.A.J. HURKENS

Abstract. We characterize adjacency of edge covers on the edge cover polytope of a graph $G=(V,E)$, and derive that the diameter of the edge cover polytope is equal to $|E|-\rho(G)$, where $\rho(G)$ is the minimum size of an edge cover.

Introduction.

Let $G=(V,E)$ be a simple graph without isolated vertices. A subset $F \subseteq E$ is called an *edge cover* of G, if every vertex of G is covered by at least one edge in F. The *edge cover polytope* of G is the convex hull (in $\mathbb{R}^E$) of the incidence vectors of the edge covers. Two edge covers are called *adjacent* if the corresponding incidence vectors are adjacent vertices of the edge cover polytope. The *diameter* of a polytope is the maximum distance between vertices of the polytope, where the *distance* between vertices u,v of a polytope, denoted by dist(u,v), is the minimum number of edges of the polytope that one must pass by on a walk on the polytope from one vertex to another along the edges. As usual, $\rho(G)$ denotes the edge cover number, i.e., the minimum size of an edge cover of G.

In this paper we characterize adjacency of edge covers, and with the help of this characterization we prove that the diameter of the edge cover polytope equals $|E|-\rho(G)$.

Remark. A well-known result of Edmonds and Johnson [1970] (which however we do not need in our proofs) implies that the edge cover polytope is described by the following linear inequalities in variable $x = (x_e | e \in E) \in \mathbb{R}^E$:

$$\begin{array}{lll} (1) & 0 \le x_e \le 1 & , e \in E \\ & \sum_{e: e\cap U \neq \emptyset} x_e \ge \lceil |U|/2 \rceil & , U \subseteq V. \end{array}$$

Determining the diameter of the edge cover polytope turns out to be more difficult than determining the diameter of the matching polytope.

For the matching polytope the characterization of adjacency, and the determination of the diameter is rather easy: Two matchings M_1 and M_2 have adjacent incidence vectors if and only if the symmetric difference $M_1 \Delta M_2$ is connected. The diameter of the matching polytope is equal to the cardinality of a maximum matching. The more general case of (1-capacitated) b-matchings is dealt with in Hurkens [1988].

Preliminaries.

Let F_1 and F_2 denote distinct edge covers of $G=(V,E)$. We define a subset S of $F_1 \Delta F_2$ to be *exchangeable*, if both $F_1 \Delta S$ and $F_2 \Delta S$ are edge covers of G. E.g., $\emptyset$ and $F_1 \Delta F_2$ are exchangeable. In order to prove our first theorem below and to understand its consequences, we need to know better what a 'minimal' nonempty exchangeable subset looks like. We define a *switch-way* T *(with respect to* F_1 *and* F_2*)* to be a sequence of vertices and edges $(v_0, e_1, v_1, \ldots, e_k, v_k)$ (with $k \geq 1$), where $e_i = \{v_{i-1}, v_i\} \in E$, for $i = 1, \ldots, k$, such that

(2) (i) $e_1, \ldots, e_k \in F_1 \Delta F_2$;

(ii) $e_i \neq e_j$, for $i \neq j$;

(iii) $|\{e_i, e_{i+1}\} \cap F_1| = 1$, $i=1, \ldots, k-1$.

We say that such a T has *length* k, and that v_0 and v_k are the *endpoints* of T, while $\{v_1, v_2, \ldots, v_{k-1}\}$ is called the *interior* of T. T is *simple*, if we have $[(v_i = v_j) \Rightarrow (i=j \text{ or } \{i,j\}=\{0,k\})]$. Define $V_0 := \{v \in V \mid \deg_{F_1 \cap F_2}(v) = 0\}$, and let $\langle V_0 \rangle$ denote the subgraph induced by V_0.

Notice, that the edge-set of a switch-way T is exchangeable, if

(3) T has at least one of the following properties:

(α) the endpoints of T are in $V \setminus V_0$;

(β) the endpoints of T coincide and T has even length ;

(γ) one endpoint of T is in $V \setminus V_0$, while the other is in the interior of T ;

(δ) both endpoints are in the interior of T .

To see this one should realize that exchanging edges along an alternating path causes no problems for vertices on the interior of this path. One only has to see to it that the endpoints remain covered. By definition of V_0, an endpoint v in $V \setminus V_0$ has the property that $\deg_{F_1 \cap F_2}(v) \geq 1$.
A switch-way satisfying (3) which has an edge-set that is inclusion-wise minimal with respect to all edge-sets of switch-ways satisfying (3) is called *independent*. Notice that an independent switch-way has all its interior vertices in V_0, has no repeating vertices (with a possible exception for the endpoints), and is of one of the following forms:

(4) (α) a simple path with endpoints in $V \setminus V_0$;
(β) a simple even circuit in $\langle V_0 \rangle$;
(γ) a simple odd circuit in $\langle V_0 \rangle$ connected to $V \setminus V_0$ by a simple path;
(δ) two simple odd circuits in $\langle V_0 \rangle$, and a simple path (possibly with length 0) connecting them .

A switch-way T that is not exchangeable must have at least one endpoint, v say, such that v is in V_0, and v is not in the interior of T. We claim

(5) T can be extended with an edge incident with v, thus yielding a larger switch-way.

Proof of claim (5).
Assume that $v=v_k$, and that $e_k \in F_1$. Then it follows that $\deg_{F_2 \setminus F_1}(v) = \deg_{F_2}(v) - \deg_{F_1 \cap F_2}(v) = \deg_{F_2}(v) \geq 1$. □

This observation leads to the following

Lemma 1.
If F_1 and F_2 are distinct edge covers, then $F_1 \Delta F_2$ contains a subset of edges that forms an independent switch-way.

Proof of Lemma 1.

Take an edge $e \in F_1 \Delta F_2$, and define $T_1 := (v_0, e, v_1)$, where v_0 and v_1 are the ends of e. By (5) we can now find a series of switch-ways $T_1, \ldots, T_k$ with the properties that:

(i) T_i is not exchangeable for $i<k$;

(ii) T_i is an extension of T_{i-1} for $i = 2,\ldots,k$, in the sense of (5);

(iii) T_k is exchangeable, and satisfies (3).

As T_k satisfies (3), it contains an independent switch-way. □

Let x_F denote the incidence vector of an edge cover F. We use the following well-known characterization of adjacency:

(6) (F_1, F_2 are adjacent edge covers) $\Leftrightarrow$

$$\left[\left[\tfrac{1}{2}\cdot(x_{F_1}+x_{F_2}) = \Sigma_F(\lambda_F x_F) \ , \ \lambda_F \geq 0 \ , \ \Sigma_F(\lambda_F)=1 \ \right] \Rightarrow \lambda_{F_1} = \lambda_{F_2} = \tfrac{1}{2} \right],$$

where F ranges over all edge covers.

Characterization of adjacency of edge covers.

Theorem 1.

Let F_1 and F_2 be distinct edge covers of $G=(V,E)$. Then the following are equivalent:

(i) F_1 and F_2 are adjacent ;

(ii) $\emptyset$ and $F_1 \Delta F_2$ are the only exchangeable subsets of $F_1 \Delta F_2$;

(iii) $F_1 \Delta F_2$ forms an independent switch-way.

Proof of Theorem 1.

(i)$\Rightarrow$(ii): Suppose F_1 and F_2 are adjacent, and let $S \subseteq F_1 \Delta F_2$ be exchangeable. Then $F_1' := F_1 \Delta S$ and $F_2' := F_2 \Delta S$ are edge covers, such that $\tfrac{1}{2}\cdot(x_{F_1}+x_{F_2}) = \tfrac{1}{2}\cdot(x_{F_1'}+x_{F_2'})$. With (6) it follows that $\{F_1, F_2\} = \{F_1', F_2'\}$ and therefore we conclude that $S=\emptyset$ or $S=F_1 \Delta F_2$.

(ii)⇒(iii): Suppose that $\emptyset$ and $F_1 \Delta F_2$ are the only exchangeable subsets of $F_1 \Delta F_2$. By Lemma 1 we know that there is an independent switch-way T contained in $F_1 \Delta F_2$. It follows that $F_1 \Delta F_2$ is the edge-set of T.

(iii)⇒(i): If $F_1 \Delta F_2$ forms an independent switch-way, then it has one of the forms mentioned in (4). It is easily verified that the right hand side of (6) holds in each of the cases (4) (α), (β), (γ) and (δ). We may therefore conclude that F_1 and F_2 are adjacent. As an example we show how to treat case (4)(γ).
Let $v_0, e_1, \ldots, e_t, v_t, e_{t+1}, v_{t+1}, \ldots, e_{t+2k+1}, v_{t+2k+1} = v_t$ denote a switch-way of type (4)(γ), with $v_0 \in V \backslash V_0$, $v_i \in V_0$, for $i \geq 1$, and all vertices distinct, except for $v_{t+2k+1} = v_t$. Assume that $e_i \in F_1$, if i is odd, and $e_i \in F_2$, if i is even. Writing x: $E \longrightarrow \{0,1\}$ as a row vector we may order E such that x can be decomposed as follows:
$x = [x|_{E \backslash (F_1 \cup F_2)} | x|_{F_1 \cap F_2} | x(e_1), x(e_2), \ldots, x(e_t) | x(e_{t+1}), \ldots, x(e_{t+2k+1})]$.
Solving $\frac{1}{2} \cdot (x_{F_1} + x_{F_2}) = [0, \ldots, 0 | 1, \ldots, 1 | \frac{1}{2}, \ldots, \frac{1}{2} | \frac{1}{2}, \ldots, \frac{1}{2}] = \sum_F (\lambda_F x_F)$, $\lambda_F \geq 0$, $\sum_F (\lambda_F) = 1$ we have to take into consideration only those F, for which x_F has an all-zero first component, an all-one second component, as third component ...,0,1,0,1 or ...,1,0,1,0 , and as fourth component 1,0,1,...,0,1 or 0,1,0,...,1,0 . This follows from the observation that each of the vertices $v_1, \ldots, v_{t-1}, v_{t+1}, \ldots, v_{t+2k}$ is covered exactly once by F_i, for i=1,2. Clearly, the combination with $x(e_t) = x(e_{t+1}) = 0$ does not yield an edge cover, as it doesnot cover v_t. So finally we have to solve $(x_{F_1} + x_{F_2})/2 = \alpha x_{F_1} + \beta x_{F_2} + \gamma [0, \ldots, 0 | 1, \ldots, 1 | \ldots, 0, 1 | 1, 0, \ldots, 1]$, with $\alpha, \beta, \gamma \geq 0$ and $\alpha + \beta + \gamma = 1$. This has one solution, viz. $\alpha = \beta = \frac{1}{2}$, $\gamma = 0$. □

Bounds on distances between edge covers.

Theorem 2.

Let F_1 and F_2 be distinct edge covers of G.
Then there exists an edge cover F_* of G, such that $F_1 \cap F_2 \subseteq F_* \subseteq F_1$ and $\text{dist}(F_1, F_2) \leq |F_1 \cup F_2| - |F_*|$.

Here $\mathrm{dist}(F_1,F_2)$ denotes the distance on the edge cover polytope between the vertices corresponding to F_1 and F_2.

Proof of Theorem 2.

We give a proof of Theorem 2 by induction on $|F_1 \Delta F_2|$.

If $|F_1 \Delta F_2|=1$, then it is clear that F_1 and F_2 are adjacent, and that the edge cover $F_*:= F_1 \cap F_2$ satisfies $\mathrm{dist}(F_1,F_2) = 1 \le |F_1 \cup F_2|-|F_*|$. So we may assume that $|F_1 \Delta F_2|>1$.

Suppose there is an edge $e \in F_1 \backslash F_2$, such that $F_1 \backslash \{e\}$ is an edge cover of G. Then define $F_1' := F_1 \backslash \{e\}$, and apply the induction hypothesis to F_1' and F_2, for which we know that $|F_1' \Delta F_2| = |F_1 \Delta F_2|-1$. We find that $\mathrm{dist}(F_1',F_2) \le |F_1' \cup F_2|-|F_*'|$, for some edge cover F_*' with $F_1' \cap F_2 \subseteq F_*' \subseteq F_1'$. Taking $F_*:=F_*'$, we are done with the proof, since then $F_1 \cap F_2 \subseteq F_1' \cap F_2 \subseteq F_* \subseteq F_1' \subseteq F_1$ and we have $\mathrm{dist}(F_1,F_2) \le \mathrm{dist}(F_1,F_1')+\mathrm{dist}(F_1',F_2) \le 1+|F_1' \cup F_2|-|F_*'| = |F_1 \cup F_2|-|F_*|$.

Hence we may assume that, from now on,

(7) For no edge $e \in F_1 \backslash F_2$, the set $F_1 \backslash \{e\}$ is an edge cover of G.

We distinguish two cases.

Case 1. F_1 and F_2 are adjacent. We can take $F_*:=F_1$, so that clearly $\mathrm{dist}(F_1,F_2) = 1 \le |F_1 \cup F_2|-|F_1| = |F_1 \cup F_2|-|F_*|$, where the inequality follows from (7) (if $F_1 \cup F_2=F_1$, then $F_2 \subseteq F_1$, whence $F_2=F_1$ by (7)).

Case 2. F_1 and F_2 are not adjacent. As a consequence of Lemma 1 and Theorem 1, $F_1 \Delta F_2$ contains an independent switch-way $S \neq F_1 \Delta F_2$. Define $F_1':=F_1 \Delta S$, and apply the induction hypothesis both to F_1, F_1' (for which we know $|F_1 \Delta F_1'| = |S| < |F_1 \Delta F_2|$) and to F_1', F_2 (for which we know $|F_1' \Delta F_2| = |F_1 \Delta F_2|-|S| < |F_1 \Delta F_2|$).

First, we find that there exists an edge cover F_*', such that $F_1 \cap F_2 \subseteq F_1 \cap F_1' \subseteq F_*' \subseteq F_1$, nd $\mathrm{dist}(F_1,F_1') \le |F_1 \cup F_1'|-|F_*'| = |F_1|+|F_2 \cap S|-|F_*'|$. From assumption (7) it follows that $F_*'=F_1$, and therefore $\mathrm{dist}(F_1,F_1') \le |F_2 \cap S|$.

Second, we have $\mathrm{dist}(F_1',F_2) \le |F_1' \cup F_2|-|F_*''| = |F_1 \cup F_2|-|F_1 \cap S|-|F_*''|$. Here F_*'' is an edge cover, such that $(F_1 \cap F_2) \cup (F_2 \cap S) = F_1' \cap F_2 \subseteq F_*'' \subseteq F_1' = (F_2 \cap S) \cup F_1 \backslash (F_1 \cap S)$. Since S satisfies (3), we have that $F_1 \cap S$ and $F_2 \cap S$ cover the same subset of vertices of V_0 $(:= \{v\in V | \deg_{F_1\cap F_2}(v)=0\})$. Therefore we know that also $F_* := F_*'' \Delta S = (F_*'' \cup (F_1\cap S))\setminus(F_2\cap S)$ is an edge cover of G, with $F_1\cap F_2 \subseteq F_* \subseteq F_1$, and $|F_*| = |F_*''|-|F_2\cap S|+|F_1\cap S|$. So $\mathrm{dist}(F_1',F_2) \leq |F_1\cup F_2|-|F_1\cap S|-|F_*''| = |F_1\cup F_2|-|F_2\cap S|-|F_*|$.

Combining the two results we find that $\mathrm{dist}(F_1,F_2) \leq \mathrm{dist}(F_1,F_1') + \mathrm{dist}(F_1',F_2) \leq |F_2\cap S| + |F_1\cup F_2|-|F_2\cap S|-|F_*| = |F_1\cup F_2|-|F_*|$. □

Theorem 3.

The diameter of the edge cover polytope of G equals $|E|-\rho(G)$, where $\rho(G)$ is the minimum size of an edge cover of G.

Proof of Theorem 3.

Let $\Delta(G)$ denote the diameter of the edge cover polytope of G.
A corollary of Theorem 1 is that if F_1 and F_2 are adjacent edge covers, then $||F_1|-|F_2|| \leq 1$, since $F_1\Delta F_2$ is a switch-way. As $|E|$ is the size of a (largest) edge cover of G, and $\rho(G)$ is the size of some minimum edge cover, say F, it is clear that $\Delta(G) \geq \mathrm{dist}(E,F) \geq |E|-|F| = |E|-\rho(G)$.

To show that $\Delta(G) \leq |E|-\rho(G)$, let F_1,F_2 be distinct edge covers of G. We know, by Theorem 2, that there is an edge cover F_* of G, such that $F_1\cap F_2 \subseteq F_* \subseteq F_1$, and $\mathrm{dist}(F_1,F_2) \leq |F_1\cup F_2|-|F_*|$. Since $|F_*| \geq \rho(G)$ we have $|F_1\cup F_2|-|F_*| \leq |E|-\rho(G)$, which concludes the proof. □

References.

[1970] J. Edmonds & E.L. Johnson, Matching, a well solved class of integer linear programs, *Combinatorial Structures and Their Applications* (R. Guy, et al., eds.), Gordon and Breach, New York, 1970, pp. 89-92.

[1988] C.A.J. Hurkens, On the Diameter of the *b*-matching Polytope, *Combinatorics* (A. Hajnal, L. Lovász and V.T. Sós, eds.), North-Holland, Amsterdam, 1988, pp. 301-307.

COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
52. COMBINATORICS, EGER (HUNGARY), 1987

On the Diameter of the b-matching Polytope

C. A. J. HURKENS *

1. Introduction

Let $G = (V, E)$ be a simple graph, and $b : V \rightarrow \mathbb{Z}_+$. A subset $M \subseteq E$ is called a *b-matching* of G, if every vertex v of G is covered by at most b_v edges in M. The *b-matching polytope* of G is the convex hull (in $\mathbb{R}^E$) of the incidence vectors of the b-matchings. Two b-matchings are called *adjacent* if the corresponding incidence vectors are adjacent vertices of the b-matching polytope. In this paper we characterize adjacency of b-matchings and derive an upper bound on the distance between two b-matchings. With the help of these results we can prove that the diameter of the b-matching polytope equals $\nu_b(G)$. Here $\nu_b(G)$ denotes the maximum size of a b-matching. The *diameter* of a polytope is the maximum distance between vertices of the polytope, where the *distance* between vertices of a polytope is the minimum number of edges of the polytope that one must pass by on a walk on the polytope from one vertex to another along the edges.

Remark. A well-known result of Edmonds and Johnson [1970] (which however we do not need in our proofs) implies that the b-matching polytope is described by the following linear inequalities in variable $x = (x_e | \; e \in E) \in \mathbb{R}^E$:

$$\begin{aligned} 0 \le x_e \le 1, &\qquad e \in E \\ \sum_{e \ni v} x_e \le b_v, &\qquad v \in V \end{aligned} \tag{1}$$

---

* Research supported by the Netherlands Organization for Scientific Research (N. W. O.), through the Stichting Mathematisch Centrum.

$$\sum_{e\subseteq U} x_e + \sum_{e\in F} x_e \leq \left\lfloor \left(\sum_{v\in U} b_v + |F| \right) /2 \right\rfloor, \; U \subseteq V, \; F \subseteq \delta(U).$$

Note, that a special case of b-matching is *matching* (take $b_v = 1$, for all v). For the matching-polytope the characterization of adjacency, and the determination of the diameter is rather easy. Two matchings M_1 and M_2 have adjacent incidence vectors if and only if the symmetric difference $M_1 \triangle M_2$ is connected (Chvátal [1975]). The diameter of this polytope is equal to the cardinality of a maximum matching.

Also, if C is an *edge cover* in a graph $G = (V, E)$, then its complement, $E \setminus C$, is a special case of b-matching (where $b_v = \deg_G(v) - 1$, for all v). The problem of adjacency and distance on the edge cover polytope has been dealt with in a previous paper (Hurkens [1986]). The diameter of the edge cover polytope is shown to be $|E| - \rho(G)$, where $\rho(G)$ is the minimum size of an edge cover in G.

It turns out, that the general case of b-matchings is much more complicated than the two special cases mentioned above. This holds in particular for the characterization of adjacency. To this purpose, we study a special type of alternating paths, which have the property that they use an edge at most once, and a vertex at most twice, with the restriction that no evensized, 'almost simple', alternating circuit is properly contained in this path. Paths of this type will be called *near-simple switch-ways*. Posing even more restrictions, such as degree constraints on the vertices of these paths, and minimality, brings us to the concept of *independent switch-ways*. We will show that two b-matchings are adjacent, only if their symmetric difference is such an independent switch-way (Theorem 1). This gives us a tool with which we can find an upper bound on the distance between two b-matchings (Theorem 2), and finally calculate the diameter of the b-matching polytope (Theorem 3).

2. Preliminaries

Let M_1 and M_2 denote distinct b-matchings of $G = (V, E)$. We define a subset $S \subseteq M_1 \triangle M_2$ to be *exchangeable*, if both $M_1 \triangle S$ and $M_2 \triangle S$ are b-matchings of G. Clearly $\emptyset$ and $M_1 \triangle M_2$ are exchangeable. In order to prove the first theorem below and to understand its consequences, we need to know better what a 'minimal' nonempty exchangeable subset looks like. We define a *switch-way* T (*with respect to* M_1 *and* M_2) to be a sequence of vertices and edges $(v_0, e_1, v_1, \ldots, e_k, v_k)$ (with $k \geq 1$), where $e_i = \{v_{i-1}, v_i\} \in E$, for $i = 1, \ldots, k$ such that

(2) $\quad$ (*i*) $\; e_1, \ldots, e_k \in M_1 \triangle M_2$;

$\quad\quad$ (*ii*) $\; e_i \neq e_j$, *for* $i \neq j$;

$\quad\quad$ (*iii*) $\; |\{e_i, e_{i+1}\} \cap M_1| = 1$, $i = 1, \ldots, k-1$.

We say that such a T has *length* k, and that v_0 and v_k are the *endpoints* of T, while $\{v_1, v_2, \ldots, v_{k-1}\}$ is called the *interior* of T. In this context we use the word *path*, if the endpoints are distinct, and *circuit*, if they coincide. T is *simple*, if we have $[(v_i = v_j \text{ and } i < j \Rightarrow (i, j = 0, k)]$. T is called *near-simple*, if

(i) $[(v_i = v_j \text{ and } i < j) \Rightarrow (i, j = 0, k \text{ or } j - i \text{ is odd})]$, and

(ii) there are no numbers $i < j < s < t$, such that $v_i = v_s$ and $v_j = v_t$.

The following figure shows, as an example, a near-simple path of even length. The black dots denote the endpoints of this path.

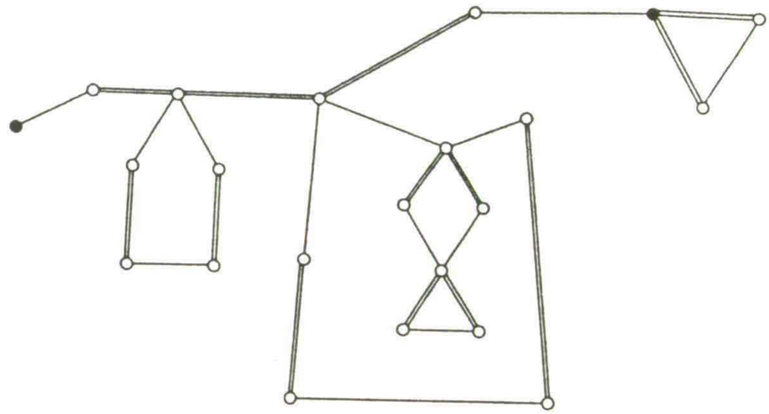


Figure 1.

A special subset of the vertices is defined by $V_+ := \{v \in V \mid \deg_{M_1 \cup M_2}(v) > b_v\}$. Notice, that the edge-set of a switch-way T is exchangeable, if

T has one of the following properties: (3)

(α) the endpoints of T are in $V \setminus V_+$;

(β) the endpoints of T coincide and T has even length;

(γ) one endpoint of T is in $V \setminus V_+$, while the other, v say, is in the interior of T, and has $\deg_{M_1 \cup M_2}(v) = b_v + 1$;

(δ) both endpoints are in the interior of T, and satisfy the degree constraint of (γ).

A switch-way satisfying (3) which has an edge-set that is inclusion-wise minimal with respect to all edge-sets of switch-ways satisfying (3) is called *independent*.

Notice that an independent switch-way has all its interior vertices in V_+, and has one of the following forms:

(4) (α) a near-simple path or circuit with endpoints in $V \setminus V_+$;

(β) a near-simple even circuit in $< V_+ >$;

(γ) a near-simple odd circuit in $< V_+ >$ connected to $V \setminus V_+$ by a near-simple path;

(δ) two near-simple odd circuits in $< V_+ >$, and a near-simple path connecting them.

A near-simple switch-way T that is not exchangeable must have at least one endpoint, v say, such that v is in V_+, and v is not in the interior of T, or has $\deg_{M_1 \cup M_2}(v) > b_v + 1$. Assume that $v = v_k$, and that $e_k \in M_1$. If v is not in the interior of T, then it follows that $\deg_{M_2 \setminus M_1}(v) \geq 1$. If v is in the interior of T, and has $\deg_{M_1 \cup M_2}(v) > b_v + 1$, then it is easily verified that $\deg_{M_2 \setminus M_1}(v) \geq 2$. We observe

(5) A non-exchangeable near-simple switch-way can be extended with an edge at one of its endpoints, thus yielding a larger switch-way.

This observation leads to the following

Lemma 1. *If M_1 and M_2 are distinct b-matchings, then $M_1 \triangle M_2$ contains a subset of edges that forms an independent switch-way.*

Proof. Take an edge $e \in M_1 \triangle M_2$, and define $T_1 := (v_0, e, v_1)$, where v_0 and v_1 are the ends of e. We can now define a series of switch-ways $T_1, \ldots, T_k$ with the properties:

(i) T_i is near-simple and not exchangeable for $i < k$;

(ii) T_i is an extension of T_{i-1} for $i = 2, \ldots, k$, in the sense of (5);

(iii) T_k is near-simple and exchangeable, or T_k is not near-simple.

It is easily verified that T_k is independent itself, or contains a smaller near-simple and exchangeable switch-way. ■

Let χ_M denote the incidence vector of a b-matching M. We use the following well-known characterization of adjacency:

(6) $(M_1, M_2 \text{ are adjacent } b\text{-matchings}) \Leftrightarrow$

$$\left[\left((\chi_{M_1} + \chi_{M_2})/2 = \sum_M (\lambda_M \chi_M), \lambda_M \geq 0, \sum_M (\lambda_M) = 1\right) \Rightarrow \lambda_{M_1} = \lambda_{M_2} = 1/2\right],$$

where M ranges over all b-matchings.

3. Characterization of adjacency of b-matching

Theorem 1. *Let M_1 and M_2 be distinct b-matchings of $G = (V, E)$. Then the following are equivalent:*

(i) M_1 and M_2 are adjacent;

(ii) $\emptyset$ and $M_1 \triangle M_2$ are the only exchangeable subsets of $M_1 \triangle M_2$;

(iii) $M_1 \triangle M_2$ forms an independent switch-way.

Proof. $(i) \Rightarrow (ii)$: Suppose M_1 and M_2 are adjacent, and let $S \subseteq M_1 \triangle M_2$ be exchangeable. Then $M_1' := M_1 \triangle S$ and $M_2' := M_2 \triangle S$ are b-matchings, such that $(\chi_{M_1} + \chi_{M_2})/2 = (\chi_{M_1'} + \chi_{M_2'})/2$. With (6) it follows that $\{M_1, M_2\} = \{M_1', M_2'\}$ and therefore we conclude that $S = \emptyset$ or $S = M_1 \triangle M_2$.

$(ii) \Rightarrow (iii)$: Suppose that $\emptyset$ and $M_1 \triangle M_2$ are the only exchangeable subsets of $M_1 \triangle M_2$. By Lemma 1 we know that $M_1 \triangle M_2$ contains an independent switch-way T. The edge-set of T is exchangeable, so it follows that T is formed by $M_1 \triangle M_2$.

$(iii) \Rightarrow (i)$: If $M_1 \triangle M_2$ is an independent switch-way it has one of the forms mentioned in (4). It is easily verified that the right hand side of (6) holds in each of the cases (4) $(\alpha), (\beta), (\gamma)$ and (δ). We may therefore conclude that M_1 and M_2 are adjacent. ■

4. Bounds on distances between b-matchings

Theorem 2. *Let M_1 and M_2 be b-matchings of G. Then there is a b-matching M_* of G, such that $M_1 \subseteq M_* \subseteq M_1 \cup M_2$ and $\operatorname{dist}(M_1, M_2) \leq |M_*| - |M_1 \cap M_2|$.*

Here $\operatorname{dist}(.,.)$ denotes the distance on the b-matching polytope between two b-matchings.

Proof. We give a proof of Theorem 2 by induction on $|M_1 \triangle M_2|$. Again, let

$$V_+ := \left\{ v \in V \mid \deg_{M_1 \cup M_2}(v) > b_v \right\}.$$

If $|M_1 \triangle M_2| = 1$, then it is clear, that M_1 and M_2 are adjacent, and that the b-matching $M_* := M_1 \cup M_2$ satisfies $\operatorname{dist}(M_1, M_2) = 1 \leq |M_*| - |M_1 \cap M_2|$.

So we may assume that $|M_1 \triangle M_2| > 1$. Suppose there is an edge $e \in M_2 \setminus M_1$, such that $M_1 \cup \{e\}$ is a b-matching of G. Then define $M_1' := M_1 \cup \{e\}$, and apply the induction hypothesis to M_1' and M_2, for which we know that $|M_1' \triangle M_2| = |M_1 \triangle M_2| - 1$. We find that $\operatorname{dist}(M_1', M_2) \leq |M_*| - |M_1' \cap M_2|$, for a b-matching M_*, with $M_1' \subseteq M_* \subseteq M_1' \cup M_2$. Notice, that also $M_1 \subseteq M_* \subseteq M_1 \cup M_2$. We have $\operatorname{dist}(M_1, M_2) \leq \operatorname{dist}(M_1, M_1') + \operatorname{dist}(M_1', M_2) \leq 1 + |M_*| - |M_1' \cap M_2| = |M_*| - |M_1 \cap M_2|$.

Hence we may assume:

(7) For no edge $e \in M_2 \setminus M_1$, the set $M_1 \cup \{e\}$ is a b-matching of G.

We distinguish two cases.

Case 1. M_1 and M_2 are adjacent. We can take $M_* = M_1$, so that clearly $\text{dist}(M_1, M_2) = 1 \leq |M_1| - |M_1 \cap M_2| = |M_*| - |M_1 \cap M_2|$, where the inequality follows from the assumption (if $M_1 = M_1 \cap M_2$, then $M_1 \subseteq M_2$, whence $M_1 = M_2$, by (7)).

Case 2. M_1 and M_2 are not adjacent. As a consequence of Lemma 1 and Theorem 1, $M_1 \triangle M_2$ contains an independent switch-way $S \neq M_1 \triangle M_2$. Define $M_1' := M_1 \triangle S$, and apply the induction hypothesis both to M_1, M_1' (for which we know that $|M_1 \triangle M_1'| = |S| < |M_1 \triangle M_2|$) and to M_1', M_2 (where we have $|M_1' \triangle M_2| = |M_1 \triangle M_2| - |S| < |M_1 \triangle M_2|$).

First, we find that there exists a b-matching M_*', such that $M_1 \subseteq M_*' \subseteq$ $\subseteq M_1 \cup M_1' \subseteq M_1 \cup M_2$, and $\text{dist}(M_1, M_1') \leq |M_*'| - |M_1 \cap M_1'| = |M_*'| -$ $-|M_1| + |M_1 \cap S|$. From our assumption (7) it follows that $M_*' = M_1$, and therefore $\text{dist}(M_1, M_1') \leq |M_1 \cap S|$.

Second, we have $\text{dist}(M_1', M_2) \leq |M_*''| - |M_1' \cap M_2| = |M_*''| - |M_1 \cap M_2| -$ $-|M_2 \cap S|$. Here M_*'' is a b-matching, such that $(M_2 \cap S) \cup M_1 \setminus (M_1 \cap S) =$ $= M_1' \subseteq M_*'' \subseteq M_1' \cup M_2 = (M_1 \cup M_2) \setminus (M_1 \cap S)$. As S is near-simple, we have that $M_1 \cap S$ and $M_2 \cap S$ cover each vertex of V_+ an equal of times, with a possible exception for the endpoints of S. If an endpoint of S, v say, is in V_+ and does not coincide with the other endpoint of S, then $\deg_{M_1 \cup M_2}(v) = b_v + 1$. We therefore know that also $M_* := M_*'' \triangle S = (M_*'' \setminus (M_2 \cap S)) \cup (M_1 \cap S)$ is a b-matching of G, with $M_1 \subseteq M_* \subseteq M_1 \cup M_2$, and $|M_*| = |M_*''| - |M_2 \cap S| + |M_1 \cap S|$. So $\text{dist}(M_1', M_2) \leq |M_*''| - |M_1 \cap M_2| - |M_2 \cap S| = |M_*| - |M_1 \cap M_2| - |M_1 \cap S|$.

Combining the two results we find that $\text{dist}(M_1, M_2) \leq \text{dist}(M_1, M_1') +$ $+\text{dist}(M_1', M_2) \leq |M_1 \cap S| + |M_*| - |M_1 \cap M_2| - |M_1 \cap S| = |M_*| - |M_1 \cap M_2|$. ■

Theorem 3. *The diameter of the b-matching polytope of equals $\nu_b(G)$, where $\nu_b(G)$ is the maximum size of a b-matching of G.*

Proof. Let $\triangle(G)$ denote the diameter of the b-matching polytope of G.

A corollary of Theorem 1 is that, if M_1 and M_2 are adjacent b-matchings, then $\big||M_1| - |M_2|\big| \leq 1$, since $M_1 \triangle M_2$ is a switch-way. As $\nu_b(G)$ is the size of a maximum b-matching of G, M_* say, and 0 is the size of the minimum b-matching $\emptyset$, it is clear that $\triangle(G) \geq \text{dist}(M_*, \emptyset) \geq |M_*| = \nu_b(G)$.

To show that $\triangle(G) \leq \nu_b(G)$, let M_1, M_2 denote distinct b-matchings of G. We know, by Theorem 2, that there is a b-matching M_* of G, such that $M_1 \subseteq M_* \subseteq$ $\subseteq M_1 \cup M_2$, and $\text{dist}(M_1, M_2) \leq |M_*| - |M_1 \cap M_2|$. Since $|M_*| \leq \nu_b(G)$ we have $|M_*| - |M_1 \cap M_2| \leq \nu_b(G)$, which concludes the proof. ■

References

[1970] J. Edmonds and E. L. Johnson, Matching, a well solved class of integer linear programs, in: *Combinatorial Structures and Their Applications* (R. Guy, et al, eds.), Gordon and Breach, New York, 1970

[1975] V. Chvátal, On certain polytopes associated with graphs, *Journal of Combinatorial Theory*, (B) **18**(1975) 138-154.

[1986] C. A. J. Hurkens, On the diameter of the edge cover polytope, *preprint*.

BLOWING UP CONVEX SETS IN THE PLANE.

C.A.J. HURKENS

Abstract. Let K be a convex set in $\mathbb{R}^2$, such that every line in $\mathbb{R}^2$ meets $K+\mathbb{Z}^2$. We prove that $\alpha K+\mathbb{Z}^2 = \mathbb{R}^2$, for $\alpha \geq 1 + \frac{2}{3}\sqrt{3}$, and that this bound is best possible, thus solving a problem of Kannan and Lovász.

Introduction.

Let K be a convex set in $\mathbb{R}^2$, such that each line in $\mathbb{R}^2$ meets $K+\mathbb{Z}^2$. Kannan and Lovász raised the question whether this implies that $2K+\mathbb{Z}^2$ covers $\mathbb{R}^2$, and proved this to be true for centrally symmetric convex bodies [1986]. In general, this however does not hold. Let τ^* denote the convex hull of $(0,0)$, $(1-\sqrt{3},1)$ and $(2-\sqrt{3},\sqrt{3}-1)$. For a counterexample we take $K = \tau^*$. Then $K+\mathbb{Z}^2$ meets every line in $\mathbb{R}^2$, as is easily verified – cf. figure 1. On the other hand, $\alpha K+\mathbb{Z}^2 \neq \mathbb{R}^2$, for $0 \leq \alpha < 1+\frac{2}{3}\sqrt{3}$, since $\frac{1}{3}(1,1+\sqrt{3})-\varepsilon(0,1) \notin \alpha K+\mathbb{Z}^2$, for ε positive and sufficiently small. Notice that $1+\frac{2}{3}\sqrt{3} \approx 2.155$.

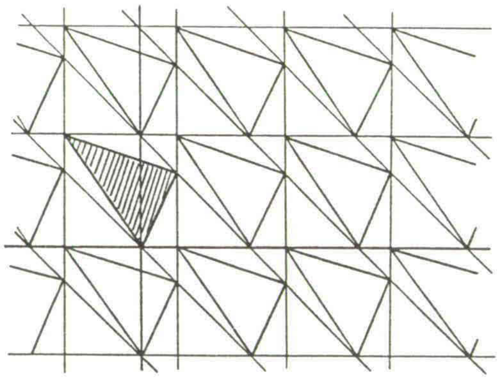


figure 1.

We will show that this is a worst case example. Furthermore it is, in some sense, unique. Let $\mathcal{T}$ denote the set of triangles in $\mathbb{R}^2$ arising from $(1+\frac{2}{3}\sqrt{3})\tau^*$ by translation and transformation with an integer matrix with determinant ± 1. We prove the following

Main theorem. Let B be a convex set such that $\mathrm{int}(B)+\mathbb{Z}^2 \neq \mathbb{R}^2$. If $\mathrm{clos}(B) \notin \mathcal{T}$, then there is a γ, $0 < \gamma < 1+\frac{2}{3}\sqrt{3}$, and a line not meeting $(\gamma^{-1})B+\mathbb{Z}^2$.

Here int(.) and clos(.) denote the interior and closure, respectively, in the usual Euclidean topology. □

Remark. The triangles in $\mathcal{T}$ are in some sense symmetric. To see this one easily constructs a one-to-one function $\Psi:\mathbb{R}^2\rightarrow\mathbb{R}^2$ mapping the lattice $\mathbb{Z}^2$ onto the lattice of regular triangles, so that for some $\delta \in \mathcal{T}$ $\Psi(\delta)$ is an equilateral triangle not containing any lattice point in its interior, and with sides making an angle of $\pi/12$ with the sides of a lattice triangle – see figure 2.

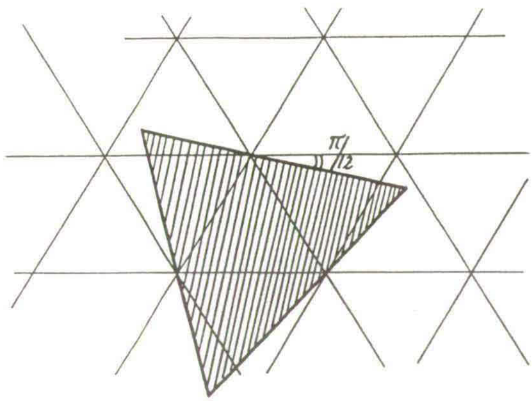


figure 2.

Before proving the main theorem we give some corollaries.

Corollary 1. Let B be a convex set such that $\mathrm{int}(B)+\mathbb{Z}^2 \neq \mathbb{R}^2$. Then for each γ, $\gamma > 1 + \frac{2}{3}\sqrt{3}$, there exists a line not meeting $(\gamma^{-1})B+\mathbb{Z}^2$. □

Corollary 2. Let $\alpha \geq 0$. Then we have $\alpha K+\mathbb{Z}^2 = \mathbb{R}^2$, for each convex set K in $\mathbb{R}^2$ such that every line in $\mathbb{R}^2$ meets $K+\mathbb{Z}^2$, if and only if $\alpha \geq 1 + \frac{2}{3}\sqrt{3}$. □

We leave it to the reader to verify that these corollaries are indeed straightforward consequences of our main theorem.

Proof of the Main Theorem.

Let B be a convex set such that $\mathrm{int}(B)+\mathbb{Z}^2 \neq \mathbb{R}^2$. Then clearly B is the translate of a convex lattice point free set. Here a set $C \subseteq \mathbb{R}^2$ is called *lattice point free* with respect to a lattice $\mathcal{L} \subseteq \mathbb{R}^2$, if we have: $\mathrm{int}(C) \cap \mathcal{L} = \emptyset$. If no lattice is specified then we take by convention the integer lattice $\mathbb{Z}^2$. For our proof we need to study only the case that B is an inclusion-wise maximal convex lattice point free set. Then it is clear that B must be closed and that one of the following holds:

[1] B is a line with an irrational slope ;

[2] B is an infinite strip with lattice points on both sides ;

[3] B is a triangle with one lattice point z_i (i=1,2,3) on each of its sides ;

[4] B is a quadrilateral with one lattice point z_i (i=1,2,3,4) on each of its sides.

For k = 1,2,3,4, we calculate $\varphi(k) := \min \{ t \mid t \geq 0$, for each B of type [k] and for each $\alpha > t$: there is a line not meeting $\alpha^{-1}B + \mathbb{Z}^2 \}$.

We claim that $\varphi(1) = 0$, $\varphi(2) = 1$, $\varphi(3) = 1 + \frac{2}{3}\sqrt{3}$, $\varphi(4) = 2$.

The first two cases are trivial and we leave it to the reader to verify them. For case [3] one should remark that:

(1) $\{z_1,z_2,z_3\}$ generates the complete lattice,

i.e., for all $z \in \mathbb{Z}^2$ there exist integers u,v,w with u+v+w=1 such that $z = uz_1+vz_2+wz_3$. This also holds for each triple taken from $\{z_1,z_2,z_3,z_4\}$ in case [4]. We use this property in the calculation of $\varphi(3)$ and $\varphi(4)$.

Proof of our claim that $\varphi(4) = 2$.

Let B be of type [4]. From observation (1) it follows that we can find a unimodular transformation mapping $\{z_1,z_2,z_3,z_4\}$ onto the set of vertices of the unit square $\{(0,1),(0,0),(1,0),(1,1)\}$. Hence we may assume that B is a quadrilateral with sides passing through (0,0), (0,1), (1,0) and (1,1). We show by means of elementary calculations that

$$(2)\qquad \min_{i=1,2}\ \max_{x \in B-B}\ e_i^T x \leq 2.$$

From this it follows directly that $\varphi(4)=2$, as there are numerous examples for which the inequality is tight. Consider the following figure:

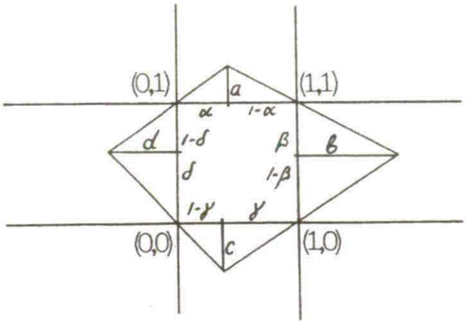


figure 3.

So $ad = \alpha(1-\delta)$, $ab = \beta(1-\alpha)$, $bc = \gamma(1-\beta)$, $cd = \delta(1-\gamma)$. Hence $1-(a+c)(b+d)$ $= 1-ab-ad-cb-cd = 1-\beta(1-\alpha)-\alpha(1-\delta)-\gamma(1-\beta)-\delta(1-\gamma) = (1-\beta-\delta)(1-\alpha-\gamma) =$ $[\alpha(1-\delta)\gamma(1-\beta)-\alpha\beta\gamma\delta][\beta(1-\alpha)\delta(1-\gamma)-\alpha\beta\gamma\delta]/\alpha\beta\gamma\delta = [abcd-\alpha\beta\gamma\delta]^2/\alpha\beta\gamma\delta \geq 0$. From $(b+d)(a+c) \leq 1$ we conclude that $\min\{a+c+1, b+d+1\} \leq 2$. □

Proof of claim $\varphi(3) = 1 + \frac{2}{3}\sqrt{3}$.

Consider three lattice points (affinely) generating the lattice, a, b and c, say. W.l.o.g. we take the centroid of the triangle $\Delta(abc)$ as the origin. Considering a,b,c as vectors in $\mathbb{R}^2$ we then have $a + b + c = 0$. Let A,B,C denote the sides of the triangle opposite to a, b and c, respectively, and let n_a, n_b, n_c denote normal vectors perpendicular to A, B and C, so that :

$$(3)\qquad \begin{aligned} -n_a^T a &= 2n_a^T b = 2n_a^T c > 0 ; \\ -n_b^T b &= 2n_b^T c = 2n_b^T a > 0 ; \\ -n_c^T c &= 2n_c^T a = 2n_c^T b > 0 . \end{aligned}$$

(See figure 4.) Then the altitudes of the triangle are $h_i := -(3/2)n_i^T i$, for $i \in \{a,b,c\}$.

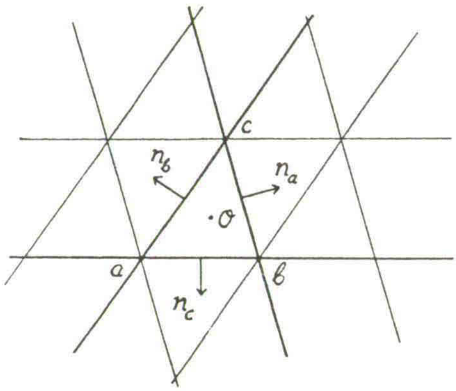


figure 4.

For a compact set K in $\mathbb{R}^2$ we define the *width* of K *with respect to* Δ(abc) as follows:

$$(4) \qquad w(K) := \min_{i\in\{a,b,c\}} \max_{\xi\in K-K} n_i^T\xi/h_i \ .$$

We will consider triangles δ in $\mathbb{R}^2$ with vertices x, y and z, opposite sides X, Y and Z, respectively, such that

(5) (i) δ has no lattice points in its interior ;
(ii) a, b and c are on the inner part of X, Y and Z, respectively.

An example is given in figure 5.

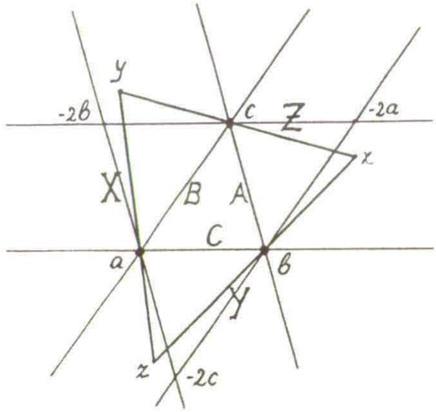


figure 5.

Our object is to show that $w(\delta) \leq 1 + \frac{2}{3}\sqrt{3}$.

The following expression says that a, b and c are on the inner parts of the sides X, Y and Z, respectively:

(6) $$\begin{bmatrix} a^T \\ b^T \\ c^T \end{bmatrix} = \begin{bmatrix} 0 & \bar\lambda & \lambda \\ \mu & 0 & \bar\mu \\ \bar\nu & \nu & 0 \end{bmatrix} \begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix}, \text{ for some } \lambda,\mu,\nu \text{ with } 0 < \lambda,\mu,\nu < 1, \text{ and } \lambda + \bar\lambda = \mu + \bar\mu = \nu + \bar\nu = 1.$$

So we can parameterize the set of triangles $\delta(xyz)$ under consideration by inverting the matrix and we find:

(7) $$\begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix} = \frac{1}{\lambda\mu\nu + \bar\lambda\bar\mu\bar\nu} \begin{bmatrix} -\bar\mu\nu & \nu\lambda & \bar\lambda\bar\mu \\ \bar\mu\bar\nu & -\bar\nu\lambda & \lambda\mu \\ \mu\nu & \bar\nu\bar\lambda & -\bar\lambda\mu \end{bmatrix} \begin{bmatrix} a^T \\ b^T \\ c^T \end{bmatrix}.$$

From the fact that none of the lattice points -2a, -2b, -2c can be in the interior of triangle δ it is easily seen that we may assume without loss of generality the following (as illustrated in figure 5) :

(8) $$n_a^T y \geq n_a^T a \geq n_a^T z \ ; \quad n_b^T z \geq n_b^T b \geq n_b^T x \ ; \quad n_c^T x \geq n_c^T c \geq n_c^T y \ .$$

[Note that this assumption affects the symmetry we had before.] These conditions can be translated in terms of λ,μ,ν and then yield :

(9) $\lambda + \mu \geq 1$; $\mu + \nu \geq 1$; $\nu + \lambda \geq 1$.

To see this note that e.g. $-n_a^T a = 2n_a^T b = 2n_a^T c > 0$, so $\left[n_a^T y \geq n_a^T a \right] \Leftrightarrow$ $\left[\bar{\mu}\bar{\nu}a^T n_a - \lambda\bar{\nu}b^T n_a + \lambda\mu c^T n_a \geq (\lambda\mu\nu+\bar{\lambda}\bar{\mu}\bar{\nu})a^T n_a \right] \Leftrightarrow \left[-2\bar{\mu}\bar{\nu}-\lambda\bar{\nu}+\lambda\mu \geq -2(\lambda\mu\nu+\bar{\lambda}\bar{\mu}\bar{\nu}) \right]$ $\Leftrightarrow\ 3\lambda(1-\bar{\mu}-\bar{\nu}) \geq 0$, whereas $n_a^T z \leq n_a^T a \Leftrightarrow 3\bar{\lambda}(1-\mu-\nu) \leq 0$. Hence $n_a^T y \geq n_a^T a \geq n_a^T z \Leftrightarrow \mu + \nu \geq 1$.

The width of triangle δ is now easily calculated and given by

$$(10) \qquad w(\delta) = \frac{\min \{\lambda,\mu,\nu\}}{\lambda\mu\nu + \bar{\lambda}\bar{\mu}\bar{\nu}} .$$

[As an example we show how to calculate $\max\limits_{\xi\in\delta-\delta} n_a^T\xi/h_a$: Clearly the maximum is attained by $\xi = x-z$, and using (7) we find $n_a^T(x-z) =$ $x^T n_a - z^T n_a = [(-\bar{\mu}\nu-\mu\nu)a^T n_a + (\nu\lambda-\bar{\nu}\bar{\lambda})b^T n_a + (\bar{\lambda}\bar{\nu}+\bar{\lambda}\nu)c^T n_a]/(\lambda\mu\nu+\bar{\lambda}\bar{\mu}\bar{\nu}) =$ $[(-\bar{\mu}\nu-\mu\nu)-\tfrac{1}{2}\cdot(\nu\lambda-\bar{\nu}\bar{\lambda})-\tfrac{1}{2}\cdot(\bar{\lambda}\bar{\mu}+\bar{\lambda}\mu)]\cdot a^T n_a/(\lambda\mu\nu+\bar{\lambda}\bar{\mu}\bar{\nu}) = -(3/2)\nu\cdot a^T n_a/(\lambda\mu\nu+\bar{\lambda}\bar{\mu}\bar{\nu})$.]

For $m := \min \{\lambda,\mu,\nu\} \neq 0$ we define $F := F(\lambda,\mu,\nu) := (w(\delta))^{-1}$. We show that $F(\lambda,\mu,\nu) \geq (1+\frac{2}{3}\sqrt{3})^{-1} = 2\sqrt{3}-3$, for $0 < \lambda,\mu,\nu \leq 1$, $\lambda+\mu \geq 1$, $\mu+\nu \geq 1$, $\nu+\lambda \geq 1$. There are two cases to be considered:

<u>Case 1</u>: $m = \lambda \leq \frac{1}{2}$. Then $\bar{\lambda} \geq \frac{1}{2} \geq \lambda$, $\mu \geq \frac{1}{2}$, $\nu \geq \frac{1}{2}$, and we have $F = [\lambda\mu\nu+\bar{\lambda}\bar{\mu}\bar{\nu}]/\lambda \geq [\lambda\mu\nu+\lambda\bar{\mu}\bar{\nu}]/\lambda = \mu\nu+\bar{\mu}\bar{\nu} = \tfrac{1}{2}\cdot(2\mu-1)(2\nu-1)+\tfrac{1}{2} \geq \tfrac{1}{2} > 2\sqrt{3}-3$;

<u>Case 2</u>: $m > \frac{1}{2}$. Define $s := \lambda+\mu+\nu$, then $s \geq 3m$ and $2(\lambda\mu+\mu\nu+\nu\lambda) - s =$ $(\lambda+\mu-1)\nu + (\mu+\nu-1)\lambda + (\nu+\lambda-1)\mu \geq (2(\lambda+\mu+\nu)-3)m = (2s-3)m$. It follows that $F = [\lambda\mu\nu + \bar{\lambda}\bar{\mu}\bar{\nu}]/m = [1-(\lambda+\mu+\nu)+(\lambda\mu+\mu\nu+\nu\lambda)]/m \geq [1-s/2+(s-3/2)m]/m =$ $[1-3m/2+(m-1/2)s]/m \geq [1-3m/2+(m-1/2)3m]/m = 3m-3+1/m \geq 2\sqrt{3} - 3$.

Notice that $F = 2\sqrt{3} - 3$ if and only if $m = \lambda = \mu = \nu = 1/\sqrt{3}$. This means that the optimum for $\max(w(\delta))$ is uniquely determined up to the choice of the sign of $n_a^T(y-a)$ (cf. assumption (8)). This remark concludes the proof of our claim that $\varphi(3) = 1 + \frac{2}{3}\sqrt{3}$. □

Using the facts that $\varphi(3) = 1 + \frac{2}{3}\sqrt{3} > 2$ and that the optimum for $\varphi(3)$ is more or less uniquely determined finally settles the proof of the main theorem. □□□

Reference.

[1986] R. Kannan and L. Lovász, *Covering Minima and Lattice Point Free Convex Bodies*, Report No. 86427 - OR, Institut für Ökonometrie und Operations Research, Bonn.

On the Existence of an Integral Potential in a Weighted Bidirected Graph

C. A. J. Hurkens*
Department of Econometrics
Tilburg University
Tilburg, The Netherlands

Dedicated to Alan J. Hoffman on the occasion of his 65th birthday.

Submitted by Alexander Schrijver

ABSTRACT

A. Schrijver proved that if A denotes the incidence matrix of a bidirected graph, and b is an integral "length" function on the edges of A, then the system $Ax \leqslant b$ has an integer solution x if and only if (i) each cycle in A has nonnegative length, and (ii) each doubly odd cycle in A has positive length. Unfortunately these cycles may be very complicated. We show that we may restrict conditions (i) and (ii) to a set of reasonably simple cycles.

1. INTRODUCTION

The purpose of this paper is to sharpen the following theorem of [2]. Let $A=(a_{ij})$ be an integral $m\times n$ matrix satisfying

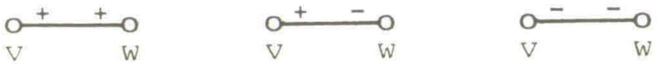
$$\sum_{j=1}^{n}|a_{ij}|=2 \qquad \text{for} \quad i=1,\dots,m, \tag{1}$$

and let b be an integral vector in $\mathbb{R}^m$. Then

the system $Ax\leqslant b$ has an integral solution x if and only if:
(i) each cycle in A has nonnegative length; (2)
(ii) each doubly odd cycle in A has positive length.

*Research supported by the Netherlands Organization for Scientific Research (NWO) through the Stichting Mathematisch Centrum.

Here the following terminology is used. With each matrix A satisfying (1) we associate a *bidirected graph*, whose *vertices* are the columns of A, and whose *edges* are the rows of A. Edge e is said to *connect* vertices v and w if $|a_{ev}| = |a_{ew}| = 1$. The sign of e at v (respectively w) is positive if a_{ev} (a_{ew}) is positive and negative if a_{ev} (a_{ew}) is negative. We can indicate this as



(3)

An edge e is said to be a *loop* at v if $|a_{ev}| = 2$, indicated as

or


(4)

A *cycle* in A is a sequence

$$(v_0, e_1, v_1, \ldots, e_d, v_d) \tag{5}$$

such that

$$\begin{array}{ll} \text{(i)} & v_0, v_1, \ldots, v_d \text{ are vertices, with } v_0 = v_d \text{, and } e_1, e_2, \ldots, e_d \\ & \text{are edges;} \\ \text{(ii)} & \text{for each } i = 1, \ldots, d \text{, either } v_{i-1} \neq v_i \text{ and } |a_{e_i v_{i-1}}| = |a_{e_i v_i}| = 1, \\ & \text{or } v_{i-1} = v_i \text{ and } |a_{e_i v_i}| = 2; \\ \text{(iii)} & \text{for each } i = 1, \ldots, d,\ a_{e_i v_i} a_{e_{i+1} v_i} < 0 \end{array} \tag{6}$$

(taking $e_{d+1} := e_1$). Examples of cycles are given by Figure 1. The *length* of a cycle (5) is, by definition,

$$\sum_{i=1}^{d} b_{e_i} \tag{7}$$

This explains condition (i) in (2).

A cycle (5) is called *doubly odd* if there exists a t with $0 < t < d$ such that

$$\begin{array}{ll} \text{(i)} & v_0 = v_t = v_d; \\ \text{(ii)} & a_{e_1 v_0} a_{e_t v_t} > 0 \text{ and } a_{e_{t+1} v_t} a_{e_d v_d} > 0; \\ \text{(iii)} & \sum_{i=1}^{t} b_{e_i} \text{ is odd and } \sum_{i=t+1}^{d} b_{e_i} \text{ is odd.} \end{array} \tag{8}$$

Conditions (i) and (ii) are illustrated by Figure 2. This explains condition (ii) of (2).

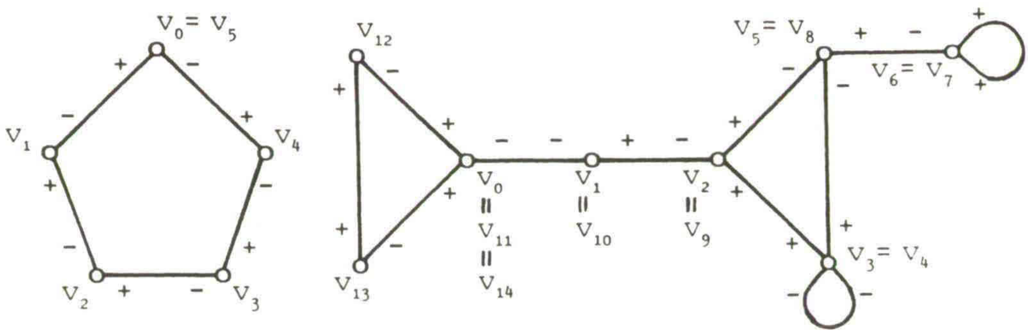


FIG. 1.

It is not sufficient to require condition (2)(i) only for *simple* cycles, i.e. those for which $v_1, \dots, v_d$ are all distinct. Similarly, condition (2)(ii) cannot be restricted to doubly odd cycles with $v_1, \dots, v_{d-1}$ all distinct.

Nevertheless, we need not verify (2) for all the cycles in A. In this paper, we describe precisely which cycles of A must be considered. In order to do this, we need the following definition. A cycle (5) is *semisimple* if there exist t and u such that

$$\begin{array}{ll} \text{(i)} & 0 \leqslant u < t - u \leqslant t < d; \\ \text{(ii)} & v_0 = v_t,\ v_1 = v_{t-1}, \dots,\ v_u = v_{t-u},\ e_1 = e_t,\ e_2 = e_{t-1}, \dots, \\ & e_u = e_{t-u+1}; \\ \text{(iii)} & v_{u+1},\ v_{u+2}, \dots, v_{d-1} \text{ are all distinct;} \\ \text{(iv)} & a_{e_1 v_0} a_{e_t v_t} > 0. \end{array} \tag{9}$$

[Condition (iv) here is superfluous if $u > 0$.]

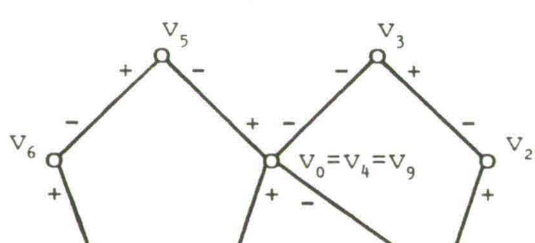


FIG. 2.

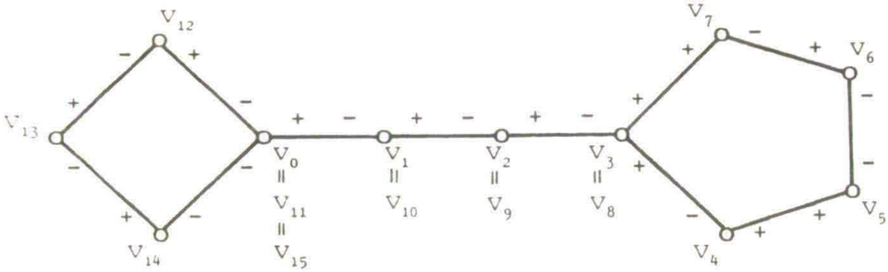


FIG. 3.

As an illustration ($u = 3$, $t = 11$, $d = 15$), consider Figure 3. Here all vertices drawn are distinct. We will show that in (2)(i) we only have to consider simple and semisimple cycles.

Characterizing the cycles necessary in (2)(ii) is more complicated. Call a sequence $P = (v_0, e_1, v_1, \ldots, e_d, v_d)$ a *path* if

$$\begin{aligned} &\text{(i)} \quad v_0, \ldots, v_d \text{ are vertices, and } e_1, \ldots, e_d \text{ are edges;} \\ &\text{(ii)} \quad \text{for } i = 1, \ldots, d, \text{ either } v_{i-1} \neq v_i \text{ and } |a_{e_i v_{i-1}}| = |a_{e_i v_i}| = 1, \\ &\qquad \text{or } v_{i-1} = v_i \text{ and } |a_{e_i v_i}| = 2; \\ &\text{(iii)} \quad \text{for } i = 1, \ldots, d-1,\ a_{e_i v_i} a_{e_{i+1} v_i} < 0. \end{aligned} \tag{10}$$

We allow the case $d = 0$. The *internal vertices* of such a path are $v_1, \ldots, v_{d-1}$. The sequence $(v_d, e_d, v_{d-1}, \ldots, e_1, v_0)$ is a path again, and is denoted by P^{-1}. If $P' = (v_0', e_1', v_1', \ldots, e_{d'}', v_{d'}')$ and $P'' = (v_0'', e_1'', v_1'', \ldots, e_{d''}'', v_{d''}'')$ are paths, we say that P' *connects* to P'' if $v_{d'}' = v_0''$ and ($a_{e_{d'}' v_{d'}'} a_{e_1'' v_0''} < 0$ or $d'd'' = 0$). If P' connects to P'' then their *concatenation* $P' \cdot P'' := (v_0', e_1', v_1', \ldots, e_{d'}', v_{d'}' = v_0'', e_1'', v_1'', \ldots, e_{d''}'', v_{d''}'')$ is again a path.

Call a cycle a *Korach cycle* if there exist paths $P_1, \ldots, P_k, Q_1, \ldots, Q_{2k}$ (with $k \geqslant 1$) such that the cycle is

$$\begin{aligned} &P_1 \cdot Q_1 \cdot P_2 \cdot Q_2 \cdot P_3 \cdots Q_{k-1} \cdot P_k \cdot Q_k \cdot P_1^{-1} \cdot Q_{k+1} \cdot P_2^{-1} \\ &\quad \cdot Q_{k+2} \cdot P_3^{-1} \cdots Q_{2k-1} \cdot P_k^{-1} \cdot Q_{2k}, \end{aligned} \tag{11}$$

where

$$\begin{aligned} &\text{(i)} \quad \text{all vertices in all } P_1, \ldots, P_k \text{ and all internal vertices in all} \\ &\qquad Q_1, \ldots, Q_{2k} \text{ are distinct;} \\ &\text{(ii)} \quad Q_i \text{ does not connect to } Q_{k+i+1}\ (i = 1, \ldots, k-1), \text{ and } Q_{k+i-1} \text{ does} \\ &\qquad \text{not connect to } Q_i\ (i = 1, \ldots, k+1). \end{aligned} \tag{12}$$

k = 2:

k = 1:

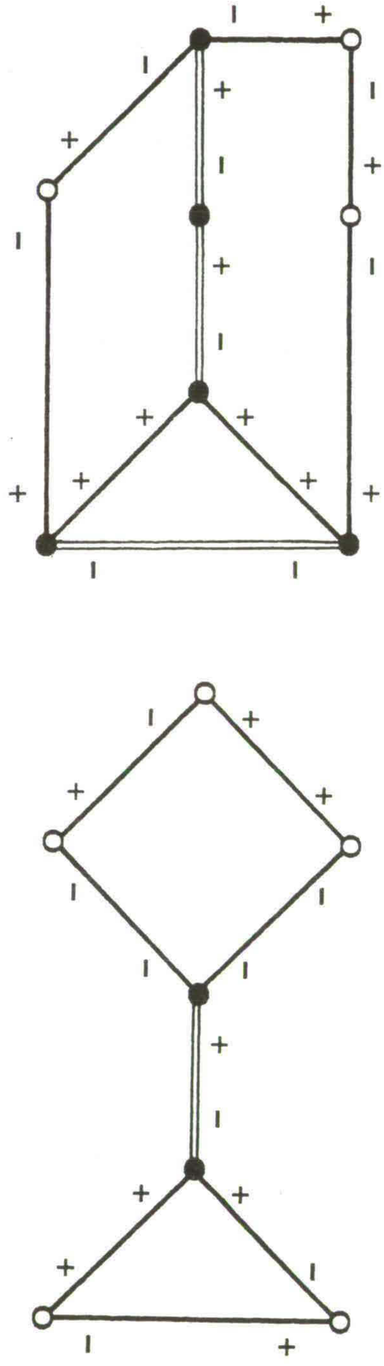
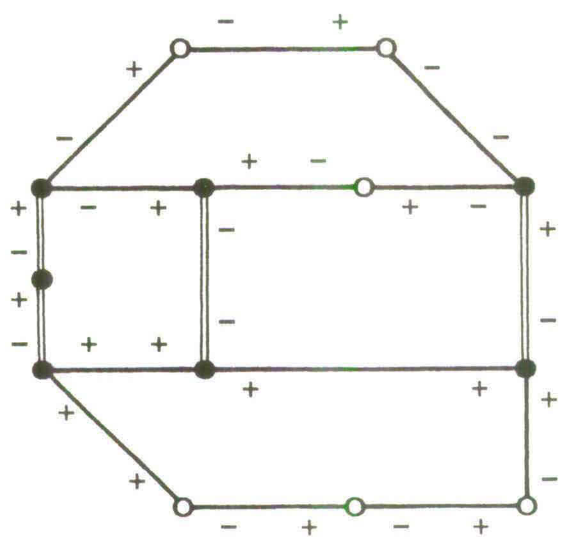


Fig. 4.

k = 3:



k = 4:

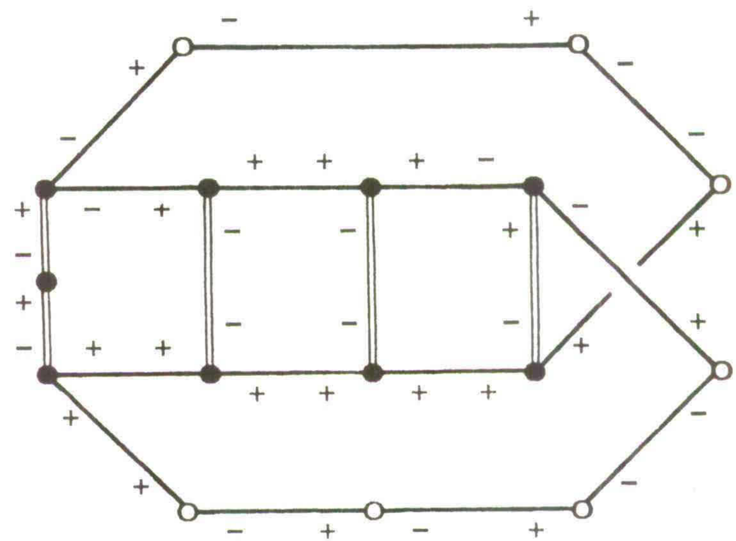


FIG. 4. (*Continued*)

$k = 5$:

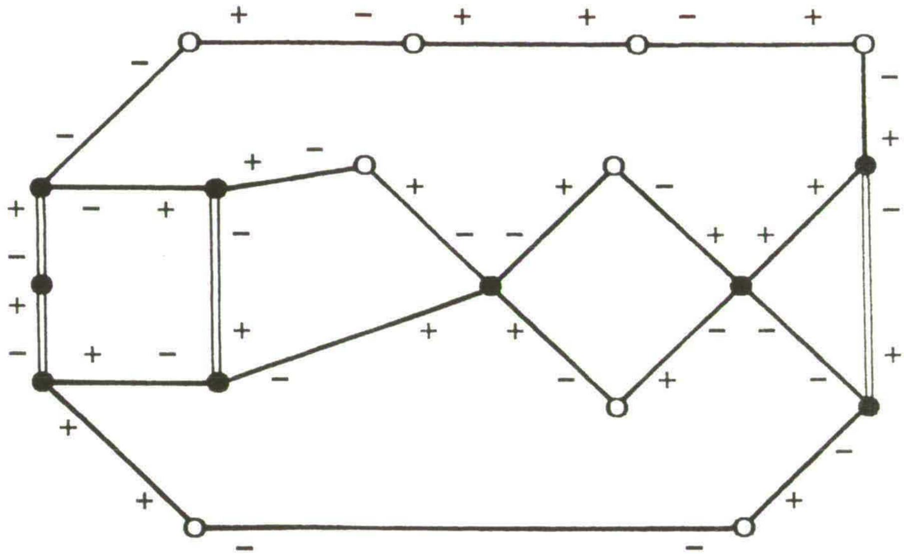


FIG. 4. *(Continued)*

[Note that (12)(i) implies that each Q_i has at least two vertices.] In Figure 4 we give examples for $k = 1, 2, 3, 4, 5$ (all vertices drawn are different, double lines and solid dots denote edges and vertices on the P_i's).
Our theorem is:

THEOREM. *The system $Ax \leqslant b$ has an integer solution x if and only if*

$$\begin{aligned} &\text{(i)} \quad \textit{each simple or semisimple cycle in A has nonnegative length;} \\ &\text{(ii)} \quad \textit{each doubly odd Korach cycle in A has positive length.} \end{aligned} \tag{13}$$

In Section 2 we give a proof of the theorem, based on a theorem of Korach [1]. In Section 3 we show that the conditions (13) cannot be reduced further: each of the cycles described is necessarily included in (13).

2. PROOF OF THE THEOREM

Our proof consists of two parts: first deriving (2)(i) from (13)(i), and second deriving (2)(ii) from (13)(i) and (ii).

I. We derive (2)(i) from (13)(i). Suppose

$$(v_0, e_1, v_1, \ldots, e_d, v_d) \tag{14}$$

is a cycle of negative length. Choose the cycle so that:

(i) the number of distinct edges occurring in (14) is as small as possible;
(ii) under condition (i), d is as small as possible. (15)

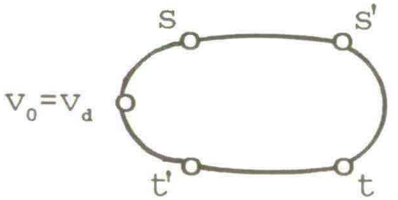
First observe that if $v_s = v_t$ for some $0 < s < t \leqslant d$, then

$$a_{e_s v_s} a_{e_t v_t} < 0. \tag{16}$$

Otherwise $(v_s, e_{s+1}, \ldots, v_t)$ and $(v_t, e_{t+1}, \ldots, v_d = v_0, e_1, \ldots, v_s)$ are cycles, at least one of them with negative length, contradicting (15).

In particular, (16) implies that no three among $v_0, v_1, \ldots, v_{d-1}$ are the same.

We derive that if $v_s = v_t$ and $v_{s'} = v_{t'}$ with $s \neq s'$ and $t \neq t'$ and $0 \leqslant s < t < d$ and $0 \leqslant s' < t' < d$, then we cannot have $s < s' < t < t'$. For suppose that



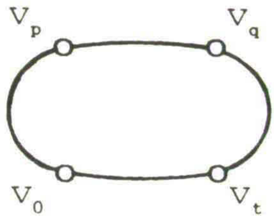
(17)

occurs. By (16), $(v_s, e_{s+1}, \ldots, e_{s'}, v_{s'} = v_{t'}, e_{t'}, \ldots, e_{t+1}, v_t)$ and $(v_{s'}, e_{s'+1}, \ldots, e_t, v_t = v_s, e_s, \ldots, e_1, v_0 = v_d, e_d, \ldots, e_{t'+1}, v_{t'})$ are cycles again, at least one of them with negative length, contradicting (15).

If (14) is not simple, then, by absence of the situation (17), we may assume without loss of generality that v_0 is chosen so that for some t, with $0 < t < d$,

$$v_0 = v_t, \text{ and for } i = t+1, \ldots, d-1,\ v_i \text{ occurs only once in } v_0, v_1, \ldots, v_{d-1}. \tag{18}$$

Then, again by absence of (17), there exist p and q, $0 \leqslant p < q \leqslant t$, such that $v_p = v_q$ and for $i = p+1, \ldots, q-1$, v_i occurs only once in $v_0, v_1, \ldots, v_{d-1}$:



(19)

(possibly $v_p = v_0$ and $v_q = v_t$). At least one of the cycles

$$(v_0, e_1, v_1, \ldots, e_q, \ v_q = v_p, e_p, \ldots, e_1, \ v_0 = v_t, e_{t+1}, \ldots, v_d)$$

and

$$(v_t, e_t, \ldots, e_{p+1}, \ v_p = v_q, e_{q+1}, \ldots, \ v_t = v_0 = v_d, e_d, \ldots, v_t)$$

has negative length and contradicts (15)(i), unless

$$(v_0, e_1, \ldots, v_p) \text{ and } (v_t, e_t, \ldots, v_q) \text{ are identical paths.} \tag{20}$$

However, if (20) holds, then (14) is semisimple.

Concluding, (14) is simple or semisimple, contradicting (13)(i).

II. We next derive (2)(ii) from (13)(i) and (ii). So by part I of this proof we may assume that (2)(i) holds. Suppose (2)(ii) does not hold. Let

$$C' = (v_0, e_1, v_1, \ldots, e_d, v_d) \tag{21}$$

be a doubly odd cycle of length 0. Let t with $0 < t < d$ satisfy (8). We show that there exists a doubly odd Korach cycle of length 0, contradicting (13)(ii). To this end, we may assume that all rows of A occur (as edges) in (21) [we can delete the rows not occurring in (21)].

In order to apply Korach's theorem [1] we construct an auxiliary undirected graph G as follows. For each vertex v of A we have two vertices v^+ and v^-. For each edge (or loop) e of A we make an edge (or loop) e^* in G, where

$$\begin{aligned}
&e^* \text{ connects } v^+ \text{ and } w^+ \text{ if } a_{ev} = a_{ew} = +1,\\
&e^* \text{ connects } v^+ \text{ and } w^- \text{ if } a_{ev} = +1 \text{ and } a_{ew} = -1,\\
&e^* \text{ connects } v^- \text{ and } w^- \text{ if } a_{ev} = a_{ew} = -1,\\
&e^* \text{ is a loop at } v^+ \text{ if } a_{ev} = +2,\\
&e^* \text{ is a loop at } v^- \text{ if } a_{ev} = -2.
\end{aligned} \tag{22}$$

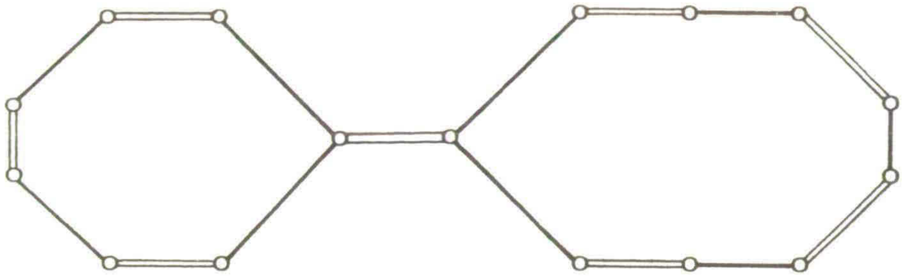


Fig. 5.

Moreover, for each vertex v of A, there is an edge in G connecting v^+ and v^-. These edges form a perfect matching M in G.

Now G does not contain a coclique (= set of pairwise nonadjacent vertices) C of size $|M|$. This follows from the facts that such a coclique C must contain exactly one vertex in every edge in M, and that the doubly odd cycle (21) gives the subgraph of G in Figure 5 (an alternating cycle with respect to M), where double lines stand for edges in M, and single lines for edges not in M. (Vertices and edges drawn different in Figure 5 may coincide.)

Now by Korach's theorem, G contains an alternating cycle of form

$$Q_1 \cdot R_1 \cdot Q_2 \cdot R_2 \ldots R_{k-1} \cdot Q_k \cdot R_k \cdot Q_1^{-1} \cdot R_{k+1} \cdot Q_2^{-1}$$
$$\cdot R_{k+2} \cdot Q_3^{-1} \ldots R_{2k-1} \cdot Q_k^{-1} \cdot R_{2k}, \tag{23}$$

where $Q_1, \ldots, Q_k$, $R_1, \ldots, R_{2k}$ are alternating paths with respect to M such that

(24)
- (i) all vertices in all $Q_1, \ldots, Q_k$ and all internal vertices in all $R_1, \ldots, R_{2k}$ are distinct;
- (ii) the first and last edges of each Q_i belong to M;
- (iii) the first and last edges of each R_i do not belong to M.

[Here, as usual, an *alternating path* with respect to M is a path $(w_0, f_1, w_1, \ldots, f_p, w_p)$, where f_i is an edge connecting w_{i-1} and w_i $(i = 1, \ldots, p)$, and where exactly one of f_{i-1} and f_i belongs to M $(i = 2, \ldots, p)$. The *internal vertices* are $w_1, \ldots, w_{p-1}$.]

In a direct way, (23) gives a Korach cycle (11) in A: replace each e^* by e, and replace each $\{v^+, v^-\}$ by v. We call this Korach cycle C''. We show that it is doubly odd and has length 0.

We first show the following:

$$\text{if} \quad y \in \mathbb{R}^m \text{ and } yA = 0 \quad \text{then} \quad yb = 0. \tag{25}$$

To see this, let for any cycle C in A, χ^C denote the incidence vector in $\mathbb{R}^m$ of C, i.e., $\chi^C(e) :=$ the number of times C passes e, for each edge (= row index) of A. By (2), the system $Ax \leqslant 2b$ has a solution x [as condition (2)(ii) is void, since $2b$ is even]. Since $\chi^{C'}b = 0$, $\chi^{C'}A = 0$, and $\chi^{C'} > 0$, we have $Ax = 2b$. This implies (25).

In particular, (25) implies, as $\chi^{C''}A = 0$, that $\chi^{C''}b = 0$, i.e., C'' has length 0.

In order to show that C'', as given by (11), is doubly odd, it suffices to show that

$$\sum_{i=2}^{k} \text{length}(P_i) + \sum_{i=1}^{k} \text{length}(Q_i) \text{ is odd.} \tag{26}$$

This fact follows from

$$\begin{array}{l} \text{if } y \text{ is an integral vector with } yA = (0,\ldots,0, \pm 2, 0,\ldots,0), \\ \text{then } yb \text{ is odd.} \end{array} \tag{27}$$

[Applying (27) to the incidence vector of $Q_1 \cdot P_2 \cdot Q_2 \ldots P_k \cdot Q_k$ gives (26).]

To see (27), let yA have its ± 2 at v_s [cf. (21)]. We may assume without loss of generality that $0 < s \leqslant t$. Let z be the incidence vector of the part

$$(v_0, e_1, v_1, \ldots, e_t, v_t) \tag{28}$$

of C'. By (8), $zA = (0,\ldots,0, \pm 2, 0,\ldots,0)$, with the ± 2 at v_t, and zb is odd, Let u be the incidence vector of the part $(v_s, e_{s+1}, v_{s+1}, \ldots, e_t, v_t)$ of C'. So $uA = (0,\ldots,0, \pm 1, 0,\ldots,0, \pm 1, 0,\ldots,0)$, with the ± 1's at v_s and v_t, or $uA = 0$ if $v_s = v_t$. Now

$$(\pm z + 2u \pm y)A = 0, \tag{29}$$

for appropriate (two) choices of $\pm$. Hence by (25), $(\pm z + 2u \pm y)b = 0$. As zb is odd and $2ub$ is even, we know that yb is odd, proving (27).

3. IRREDUNDANCY OF THE CONDITION (13)

We finally show that the condition (13) cannot be reduced any further. Call two cycles *equivalent* if they are the same up to the choice of the starting point, up to the orientation, and, in case they are semisimple, up to replacing $C_1 \cdot C_2$ by $C_1 \cdot C_2^{-1}$. We claim:

> Let A be any bidirected graph. Let C be any simple, semisimple, or Korach cycle of A. Then for some right-hand side b, C is the only cycle of A (up to equivalence) which does not satisfy the condition (13). (30)

To prove our claim, let $C = (v_0, e_1, v_1, \ldots, e_d, v_d)$.

First, let C be a simple or semisimple cycle. Define $b(e) := -2$ for each edge $e \in C$, and $b(e) := 4d + 2$ for each edge $e \notin C$. Then C has negative length. Moreover, each simple or semisimple cycle in A that contains an edge not occurring in C has positive length. We can therefore assume that all edges of A occur in C. Clearly, C is the only simple or semisimple cycle contained in A (up to equivalence), and we should consider C in (13)(i).

Next, let

$$C = P_1 \cdot Q_1 \cdot P_2 \cdot Q_2 \ldots P_k \cdot Q_k \cdot P_1^{-1} \cdot Q_{k+1} \cdot P_2^{-1} \cdot Q_{k+2} \ldots P_k^{-1} \cdot Q_{2k} \qquad (31)$$

be a Korach cycle, as in (11). Let $b(e) := (\frac{1}{2}A\mathbf{1})_e$ for $e \in C$, and $b(e) := 2d + 1$ for $e \notin C$. Here $\mathbf{1}$ denotes the all-one vector in $\mathbb{R}^n$. One easily checks that now C is doubly odd and has length 0. So $Ax \leqslant b$ has no integer solution. On the other hand, $Ax \leqslant b$ has a rational solution, viz. $x = \frac{1}{2}\mathbf{1}$. So all cycles of A have nonnegative length. Each Korach cycle of A that contains an edge not occurring in C has positive length. Hence we may assume that all edges of A occur in C.

We show that C is the only Korach cycle contained in A (up to equivalence). Define $P_{k+j} := P_j^{-1}$ for $j = 1, \ldots, k$. Let C' be a Korach cycle contained in A. Without loss of generality, we may assume that Q_1 is part of $C' \cdot C'$. Hence also $P_1 \cdot Q_1$ is part of $C' \cdot C'$. Let q be the largest number with $q \leqslant 2k$ such that

$$P_1 \cdot Q_1 \cdot P_2 \cdot Q_2 \ldots P_q \cdot Q_q \qquad (32)$$

is part of $C'\cdot C'$. If $q=2k$, then C' is equivalent to C. This follows from the following fact, which is easy to derive from the definition of a Korach cycle:

$$\text{if } \alpha \text{ is a cycle and is a proper subsequence of } \beta\cdot\beta, \text{ where } \beta \text{ is a Korach cycle, then } \alpha \text{ is equivalent to } \beta. \tag{33}$$

This should be applied to α being cycle (32), and β being C and C'.

So we may assume that $q<2k$. The part (32) of $C'\cdot C'$ must be followed by P_{q+1}. By the maximality of q, this P_{q+1} cannot be followed by Q_{q+1}, and hence it must be followed by Q_{k+q}^{-1} (taking indices modulo $2k$). That is,

$$P_1\cdot Q_1\cdot P_2\cdot Q_2\dots P_q\cdot Q_q\cdot P_{q+1}\cdot Q_{k+q}^{-1} \tag{34}$$

is part of $C'\cdot C'$. Since $P_q\cdot Q_q\cdot P_{q+1}\cdot Q_{k+q}^{-1}$ is a cycle, by (33) it is equivalent to C'. Hence $q=1$ and $k=1$, and therefore C is equivalent to C'.

REFERENCES

1 E. Korach, Packing of T-cuts, and Other Aspects of Dual Integrality, Ph.D. Thesis, Waterloo, Ontario, 1982.

2 A. Schrijver, Disjoint Circuits of Prescribed Homotopies in a Graph on a Compact Surface, Report OS-R8812, Centre for Mathematics and Computing Science, Amsterdam, The Netherlands.

Received 14 July 1988; final manuscript accepted 16 September 1988

REDUCTION OF CUT-CONDITIONS ON COMPACT SURFACES.

C.A.J. HURKENS

Abstract. - We prove that a certain cut-condition for graphs on a compact surface can be restricted to "simple" and "semi-simple" cuts. This applies to theorems of Schrijver on disjoint circuits of prescribed homotopies in a graph on a compact surface.

SECTION 1 - INTRODUCTION

Our main result is the following 'reduction theorem':

Theorem 1. *Let* S *be a compact surface and let* $G = (V(G),E(G))$ *be a graph embedded on* S. *Let* $C_1,\dots,C_k$ *be pairwise disjoint closed curves on* S, *each of them simple and orientation preserving.*

If there exists a closed curve D *on* S *satisfying :*

(1) (i) $D \cap V(G) = \emptyset$;

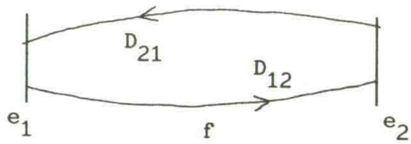
(ii) $cr(G,D) < \Sigma_i \mathrm{mincr}(C_i,D)$,

then there exists one satisfying the following further conditions :

(2) (i) D *is simple ;*

(ii) *the intersection-sequence* $\bar{D}$ *of* D *is simple or semi-simple ;*

(iii) *if* e_1,e_2 *are edges of* G, *and* f *is a face of* G, *so that part* D_{12} *of* D *traverses* f, *going from* e_1 *to* e_2, *and part* D_{21} *traverses* f, *going from* e_2 *to* e_1, *as in :*



then e_1, D_{12}, e_2 *and* D_{21} *enclose a simply connected part of face* f.

Here we use the following conventions, terminology and notation. A graph is said to be *embedded on* S, if it is embedded so that edges do not intersect. We identify a graph with its image on S. The *faces* of a graph are the components of $S \setminus G$.

A *closed curve* on S is a continuous function $C: S_1 \to S$, where S_1 denotes the unit circle $\{z \in \mathbb{C} \mid |z| = 1\}$ in the complex plane. It is *simple* if it is one-to-one. Two closed curves C and C' are called *homotopic* (*on* S), denoted by $C \sim C'$, if there exists a continuous function $\Phi: S_1 \times [0,1] \to S$ such that $\Phi(z,0) = C(z)$ and $\Phi(z,1) = C'(z)$ for all $z \in S_1$. A closed curve *is called null-homotopic* if it is homotopic to some constant function.

A closed curve is called *orientation preserving* if the notions of left and right do not change after making one orbit along the curve. In other words – if the compact surface is represented by the 2-sphere by adding a finite number of handles and a finite number of so-called 'cross-caps', then such a curve should traverse cross-caps an even number of times.

For a graph G embedded on S and a closed curve D on S we denote:

(3) $\quad \mathrm{cr}(G,D) := |\{z \in S_1 \mid D(z) \text{ belongs to } G\}|.$

Likewise, if C and D are closed curves on S:

(4) $\quad \mathrm{cr}(C,D) := |\{(y,z) \in S_1 \times S_1 \mid C(y) = D(z)\}|,$
$\quad \mathrm{mincr}(C,D) := \min \{\mathrm{cr}(\tilde{C},\tilde{D}) \mid \tilde{C} \sim C,\ \tilde{D} \sim D\}.$

It is well-known that mincr(C,D) is finite.

In order to describe self-intersections and self-crossings of a closed curve C we use:

(5) $\quad \mathrm{cr}(C) := \frac{1}{2} \cdot |\{(y,z) \in S_1 \times S_1 \mid C(y) = C(z),\ y \neq z\}|,$
$\quad \mathrm{mincr}(C) := \min \{\mathrm{cr}(\tilde{C}) \mid \tilde{C} \sim C\}.$

The notion of *self-crossing* as opposed to *self-intersection* is intuitively clear.

Let $G = (V(G),E(G))$ be a graph embedded on S and let D be a closed curve on S, with $cr(G,D) < \infty$. The *intersection-sequence* $\bar{D}$ of D is the sequence defined by:

(6) $$\bar{D} = (f_0,f_1,\ldots,f_t)$$

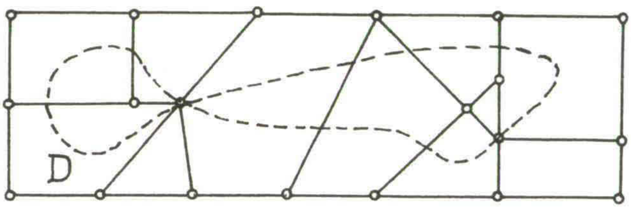
with $f_0=f_t$, where f_i is a face, edge or vertex of G, so that (6) gives the order in which D passes faces, edges and vertices of G. The number $t = 2\cdot cr(G,D)$ may be zero. Sequence (6) is called *simple* if the f_i, $i=1,\ldots,t$, are distinct. It is *semi-simple* if there are indices q,r,s with $0 \leq q \leq r \leq s \leq t$ so that:

(7) (i) $(f_0,f_1,\ldots,f_q) = (f_s,f_{s-1},\ldots,f_r)$;
(ii) $f_{q+1},f_{q+2},\ldots,f_{t-1}$ are all distinct

(so $q=s-r$).

Similarly any cyclic permutation of such a sequence will be called semi-simple. As an illustration ($q=3,r=11,s=14$) consider the following curve D:

(8)



Here all faces, edges and vertices drawn are distinct. □

A first application of the reduction theorem concerns a special case of the following 'homotopic circulation theorem' of Schrijver on the existence of fractional circulations of prescribed homotopy. We use notation as follows. Let $G=(V,E)$ be embedded on a surface S. Let $C = (v_0,e_1,v_1,\ldots, e_\ell,v_\ell)$ be a cycle in G. Then the function $\chi^C:E\to\mathbb{Z}_+$ is defined by:

(9) $$\chi^C(e) = |\{j=1,\ldots,\ell \mid e=e_j\}| \qquad \text{for } e \in E.$$

The formulation of the 'homotopic circulation theorem' is as follows:

Theorem 2. (Schrijver,[7].) *Let* $G = (V,E)$ *be a graph embedded on a compact orientable surface* S, *and let* $C_1,\dots,C_k$ *be cycles in* G. *Then there exist cycles* $C_{11},\dots,C_{1t_1}, C_{21},\dots, C_{k1},\dots,C_{kt_k}$ *in* G *and scalars* $\lambda_{11},\dots,\lambda_{kt_k} \geq 0$ *so that :*

$$(10)\qquad \text{(i)}\quad \sum_{j=1}^{t_i} \lambda_{ij} = 1 \qquad (i=1,\dots,k),$$

$$\text{(ii)}\quad \sum_{i,j} \lambda_{ij}\cdot\chi^{C_{ij}}(e) \leq 1 \qquad (e\in E),$$

if and only if for each closed curve D *on* S *not intersecting* V *we have :*

$$(11)\qquad \sum_{i=1}^{k} \operatorname{mincr}(C_i,D) \leq \operatorname{cr}(G,D).$$

□

The following corollary of our reduction theorem shows that condition (11) can be sharpened in some cases:

Theorem 3. *Let* $G = (V,E)$ *be a graph embedded on a compact orientable surface* S. *Let* $C_1,\dots,C_k$ *be pairwise disjoint and simple closed curves on* S. *If there exists a closed curve* D *on* S *not intersecting* V *for which:*

$$(12)\qquad \sum_{i=1}^{k} \operatorname{mincr}(C_i,D) > \operatorname{cr}(G,D),$$

then there exists one satisfying moreover the 'simplicity conditions' (2).

□

As a second application of the reduction theorem we give a reduction of part of the set of necessary and sufficient conditions for the existence of pairwise vertex-disjoint simple circuits, of prescribed homotopies, in a graph embedded on a compact surface. These conditions form the core of yet another theorem of Schrijver [8], and put restrictions on the set of 'dual curves' on the surface. The theorem reads as follows:

Theorem 4. (Schrijver,[8].) *Let* $G=(V,E)$ *be a graph embedded on a compact surface* S, *and let* $C_1,\dots,C_k$ *be non-null-homotopic closed curves on* S. *Then there exist pairwise vertex-disjoint simple closed curves* $\tilde{C}_1,\dots,\tilde{C}_k$ *in* G *so that* $\tilde{C}_i$ *is homotopic to* C_i, *for* $i=1,\dots,k$, *if and only if* :

(13) (i) *there exist pairwise disjoint simple closed curves* $\tilde{C}_1,\dots,\tilde{C}_k$ *on* S *so that* C_i *is homotopic to* $\tilde{C}_i$, *for* $i=1,\dots,k$;

(ii) *for each closed curve* $D:S\to S_1$: $cr(G,D) \geq \Sigma_{i=1}^{k} mincr(C_i,D)$;

(iii) *for each doubly odd closed curve* $D = D_1 \# D_2 : S_1 \to S$ *with* $D_1(1) = D_2(1) \notin G$: $cr(G,D) > \Sigma_{i=1}^{k} mincr(C_i,D)$.

□

We will not explain the precise meaning of (13)(iii) as we are concerned only with the first two conditions. Our application asserts that if condition (13)(i) is satisfied, and (13)(ii) is violated, then it is violated by a rather 'simple' curve D:

Theorem 5. *Let* S *be a compact surface. Let* $G = (V(G),E(G))$ *be a graph embedded on* S. *Let* $C_1,\dots,C_k:S_1\to S$ *be simple pairwise disjoint closed curves on* S, *each of them non-null-homotopic. If there exists a closed curve* $D:S_1\to S$, *with the property that*

(14) $$cr(G,D) < \Sigma_{i=1}^{k} mincr(C_i,D),$$

then there exists one satisfying the following further conditions :

(15) (i) D *intersects* G *only in* $V(G)$;

(ii) D *has no self-crossings* ;

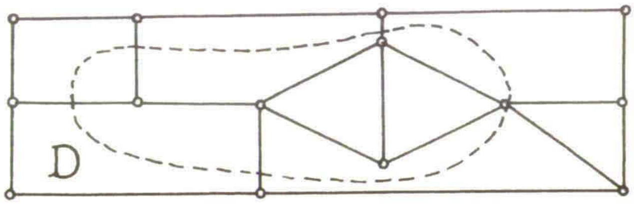
(iii) *the intersection-sequence* $\overline{D}$ *of* D *is simple or semi-simple* ;

(iv) *if* v_1,v_2 *are vertices of* G, *and* f *is a face of* G, *so that part* D_{12} *of* D *traverses* f, *going from* v_1 *to* v_2, *and part* D_{21} *traverses* f, *going from* v_2 *to* v_1, *then* $D_{12} \sim (D_{21})^{-1}$.

□

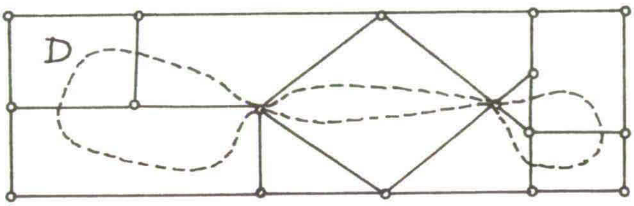
It means that if some dual cut D violates condition (13)(ii), then we can assume that both D and $\bar{D}$ are simple, as in:

(16)



or $\bar{D}$ is semi-simple, and D has no self-crossings, as in:

(17)



In both examples, all faces and all vertices drawn are distinct.

Remark. The proof of the Reduction Theorem that we will give below allows an extension asserting that if G has no component planarly embedded on S, then we can make an additional assumption, viz.:

(18) with at most two exceptions, $f \backslash D[S_1]$ is simply connected, for each face f of G traversed by D.

We will not prove this assertion, to avoid further complications in our proof.

In a sequel to this paper [4] we will give a reduction of the doubly odd closed curves, mentioned in (13)(iii), to so-called 'Korach-curves'. In Hurkens [3] a theorem is presented, providing analogous results on reducing necessary and sufficient conditions for the existence of integral solution to a matrix-inequality of a certain type.

The rest of this paper is organized as follows. In section 2 we give a brief survey of the topology of surfaces and we present some more definitions on curves and crossings of curves. We actually borrow the terminology of Schrijver (cf. [7],[8]) in order to keep our paper self-contained. So the main part of section 2 defines the language used to describe the proof of our reduction theorem. Furthermore we derive some preliminary results turning out to be useful in proving our theorem.

The proof of our reduction theorem is given in section 3. Finally in section 4 we show that theorems 3 and 5 are indeed corollaries of the reduction theorem.

SECTION 2 - PRELIMINARIES ON SURFACES AND CURVES

Surfaces and curves.

A *surface* is any arc-connected Hausdorff space S in which each point x has a neighborhood N_x homeomorphic to the complex plane $\mathbb{C}$. A surface S is *orientable* if each N_x can be oriented so that if two neighborhoods N_x and N_y intersect, then their orientations coincide on the intersection. Otherwise, S is called *non-orientable.*

Dehn and Heegaard [2] classified all compact surfaces to those spaces obtained from the 2-dimensional sphere by adding some finite number of 'handles' and some finite number of 'cross-caps'. Beside the compact surfaces we will consider in our proof only three other, non-compact surfaces, viz.

(19) - the complex plane $\mathbb{C}$;
- the annulus;
- the Möbius strip.

The *annulus* arises from $\mathbb{R}\times[0,1]$ by identifying $(x,0)$ and $(x,1)$, for each $x\in\mathbb{R}$. The *Möbius strip* arises from $\mathbb{R}\times[0,1]$ by identifying $(x,0)$ and $(-x,1)$, for each $x\in\mathbb{R}$.

Curves and paths.

A *closed curve* D on a surface S is a continuous function $D:S_1\rightarrow S$, where $S_1:=\{z\in\mathbb{C}\,|\,|z|=1\}$. An *open curve* on S is a continuous function $D:\mathbb{R}\rightarrow S$.

A *path* P on S is a continuous function $P:[0,1]\rightarrow S$. The path is said to *go from* P(0) *to* P(1), which two points are called the *end points* of P. If D is a closed curve on S, then path(D) is the path on S, defined by:

(20) $$\text{path}(D)(x) := D(e^{2\pi ix}), \quad \text{for } x \in [0,1].$$

A closed curve D on S is called *orientation-preserving,* if its orientation does not 'flip' after making one turn. A closed curve D for which the orientation does flip is called *orientation-reversing.*

If $D_1, D_2: S_1 \to S$ are closed curves with $D_1(1)=D_2(1)$, then $D_1 \# D_2$ is the closed curve with:

(21) $$(D_1 \# D_2)(z) := \begin{cases} D_1(z^2), & \text{if } \operatorname{Im} z \geq 0; \\ D_2(z^2), & \text{if } \operatorname{Im} z < 0. \end{cases}$$

Similarly $D_1 \# \ldots \# D_n$ is defined. For $n \in \mathbb{Z}$, if $D: S_1 \to S$ is a closed curve, then D^n is the closed curve with $D^n(z) := D(z^n)$ for $z \in S_1$.

Similarly, if $P_1, P_2: [0,1] \to S$ are paths with $P_1(1)=P_2(0)$, then $P_1 \# P_2$ is the path with

(22) $$(P_1 \# P_2)(x) := \begin{cases} P_1(2x) & , \text{if } 0 \leq x \leq \frac{1}{2}; \\ P_2(2x-1), & \text{if } \frac{1}{2} < x \leq 1. \end{cases}$$

If $P: [0,1] \to S$ is a path on S, then P^{-1} is the path on S defined by $P^{-1}(x) := P(1-x)$, for $x \in [0,1]$.

Homotopy.

Two closed curves D and D' are called *homotopic (on* S), which is denoted by $D \sim D'$, if there exists a continuous function $\Phi: S_1 \times [0,1] \to S$ such that $\Phi(z,0)=D(z)$ and $\Phi(z,1)=D'(z)$ for all $z \in S_1$. This defines an equivalence relation between closed curves; the class containing curve D is denoted by hom(D). A closed curve is called *null-homotopic* if it is homotopic to some constant function.

Similarly, two paths $P, P': [0,1] \to S$ are said to be *homotopic*, denoted by $P \sim P'$, if there exists a continuous function $\Phi: [0,1] \times [0,1] \to S$ so that $\Phi(x,0)=P(x)$, $\Phi(x,1)=P'(x)$, $\Phi(0,x)=P(0)$ and $\Phi(1,x)=P(1)$, for all $x \in [0,1]$. It follows that $P(0)=P'(0)$ and $P(1)=P'(1)$. Again, homotopy of paths defines an equivalence relation; the class containing path P is denoted by hom(P). A path is called *null-homotopic* if it is homotopic to a constant function.

Moreover, for $p,q \in S$, define $\operatorname{Hom}(p,q) := \{\operatorname{hom}(P) \mid P \text{ is a path from } p \text{ to } q\}$. If $p,q,r \in S$ and $\lambda \in \operatorname{Hom}(p,q)$ and $\mu \in \operatorname{Hom}(q,r)$, then define $\lambda \# \mu := \operatorname{hom}(P \# Q)$, for some arbitrary $P \in \lambda$ and $Q \in \mu$ ($\operatorname{hom}(P \# Q)$ is easily seen to be independent of the choice of $P \in \lambda, Q \in \mu$). This operation makes

Hom(p,p) to a group, the *fundamental group* $\pi_1(S)$ *of* S (as a group it is independent of p). Let id denote the identity element of Hom(p,p).

By a theorem of Poincaré [6], the 2-sphere and the complex plane are the only surfaces with trivial fundamental group. By a theorem of von Kerékjártó [5], the annulus and the Möbius strip are the only surfaces with infinite cyclic fundamental group.

The fundamental groups of the compact surfaces are well-described, and it follows that:

(23) if S is a compact surface not equal to the projective plane, and C is a closed curve on S with C^n null-homotopic ($n\geq 2$), then C itself is null-homotopic.

Covering surfaces.

A *covering surface* of a surface S is a surface S' provided with a map $\pi:S'\to S$ so that: each point p of S has a neighborhood $N \cong \mathbb{C}$ so that for each component K of $\pi^{-1}[N]$ one has that $\pi|_K$ is a homeomorphism from K onto N. π is called the *projection function* of the covering surface.

The universal covering surface.

Choose $p\in S$. Then *the universal covering surface* S' of S (*with respect to* p) is the space with point set: $\{\lambda \mid q\in S,\ \lambda\in \mathrm{Hom}(p,q)\}$, while a subset T of S' is open iff for each $\lambda\in T$, $\lambda\in\mathrm{Hom}(p,q)$, there is a neighborhood N of q in S so that N is homeomorphic to $\mathbb{C}$ and so that for each r in N and for each path P in N from q to r the point $\lambda\#\mathrm{hom}(P)$ belongs to T. The projection $\pi:S'\to S$ is the continuous function defined by $\pi(\lambda):=q$, for $\lambda\in\mathrm{Hom}(p,q)$.

For any closed curve $D:S_1\to S$ with $D(1)=q$, say, and for any $\lambda\in\mathrm{Hom}(p,q)$, the *lifting of* D *to* S' *by* λ is the open curve $D':\mathbb{R}\to S'$ defined by:

(24) $$D'(x) := \lambda\#\mathrm{hom}(D(e^{2\pi ixy})_{y\in[0,1]}) \quad \text{for } x \text{ in } \mathbb{R}.$$

In fact, D' is the unique open curve $D':\mathbb{R}\to S'$ satisfying $\pi\circ D'(x)=D(e^{2\pi ix})$ for all x in $\mathbb{R}$.

Similarly, for any path $D:[0,1]\to S$ with $D(0)=q$, say, and for any $\lambda\in Hom(p,q)$, the *lifting of* D *to* S' *by* λ is the path $D':[0,1]\to S'$ defined by: $D'(x) := \lambda \# hom(D(xy)_{y\in[0,1]})$ for $x\in[0,1]$. So $\pi\circ D'=D$.

Note the symmetry of the universal covering surface: the universal covering surface and the liftings are essentially independent of the choice of the point p. In fact, one has the following helpful result:

(25) if S is a compact surface, not equal to the 2-sphere or the projective plane, and S' is the universal covering surface of S, then S' is homeomorphic to the complex plane $\mathbb{C}$.

This was shown by Schwarz and by Poincaré [6]. It means that copies of the 'fundamental polygon' of any compact surface, except the 2-sphere and the projective plane, can be stuck together so as to form a tesselation of a space homeomorhic to $\mathbb{C}$.

The covering surface generated by a curve.

The above mentioned concept of lifting paths and curves to the universal covering surface also applies for general covering surfaces. A special kind of covering surface we will use is the surface *generated by* a curve on our surface S. It arises from 'rolling up' the universal covering surface along the lifting of a particular curve of our choice.

Let $D:S_1\to S$ be a non-null-homotopic closed curve and let $p:=D(1)$. The *covering surface generated by* D is the quotient space of the universal covering surface S' with respect to p, obtained by identifying $\lambda\in S'$ and $\mu\in S'$ iff $\lambda = hom(path(D^n))\#\mu$ for some $n\in\mathbb{Z}$. So any point of S" can be described by $\langle\lambda\rangle$ where $\lambda\in S'$ and where $\langle\lambda\rangle$ denotes the class of λ under the equivalence just defined.

Let $\pi':S'\to S''$ denote the quotient map. Then $\pi'(\lambda) = \langle\lambda\rangle$. The projection $\pi'':S''\to S$ is the function given by $\pi''(\langle\lambda\rangle) := q$, if $\lambda \in Hom(p,q)$. So $\pi''\circ\pi' = \pi$, or in diagram:

(26)
$$\begin{tikzcd} S' \arrow[rr, "\pi"] \arrow[dr, "\pi'"'] & & S \\ & S'' \arrow[ur, "\pi''"'] & \end{tikzcd}$$

If S is not the projective plane then S" has a fundamental group isomorphic to the infinite cyclic group. Hence, topologically, S" is homeomorphic to the annulus or the Möbius strip (by von Kerékjártó's classification theorem, [5]), depending on whether D is orientation-preserving or not.

Preliminary results.

Next we give some results which turn out to be useful in the proof of our theorem. The first one asserts that if we have a curve homotopic to C^n, for some orientation-preserving closed curve C, then we can 'split off' a subcurve homotopic to C. To shorten the proof we formulate a somewhat weaker statement:

(27) Let C be a non-null-homotopic orientation-preserving closed curve on S, and let D be a closed curve, $D \sim C^n$, for some $n \geq 1$, with $cr(D) < \infty$.
Then there exists a homeomorphism $\varphi: S_1 \to S_1$ homotopic to the identity map w.r.t. the surface $\mathbb{C}\setminus\{0\}$, and there exist curves $E \sim C$ and $F \sim C^{n-1}$, so that $D\circ\varphi = E\# F$.

Proof. By induction on cr(D). Let S" denote the surface generated by C. Then S" is homeomorphic to the annulus. The lifting D" of D to S" is a closed curve, and n is equal to the net number of times D" goes around the annulus. Moreover $cr(D") \leq cr(D)$.
If $cr(D") = 0$, then $n = 1$ and we can take $E = D$.
So assume $cr(D") > 0$. Without loss of generality D" intersects in $D"(1) = D"(-1)$ and we can split $D" = D_1"\# D_2"$. Let D_1 and D_2 denote the respective projections on S. Then $D = D_1\# D_2$ and $cr(D_1)+cr(D_2) < cr(D)$. Suppose $D_1"$ goes around the annulus p times, and $D_2"$ goes around q times. That is, $D_1" \sim M^p$ and $D_2" \sim M^q$, where $M: S_1 \to S"$ denotes the mid-circle of the annulus, i.e. the simple closed curve on the annulus given by $M(e^{2\pi ix}) := (0,x)$, for $x\in[0,1]$. Then $D_1\sim C^p$, $D_2\sim C^q$, and we have $p+q = n$, where p or q may be negative. Suppose that our choice of the point of self-intersection of D" is so, that $|p|+|q|$ is minimal. If both p and q are non-negative, then the proof is finished by applying the induction hypothesis to D_1 and D_2.

On the other hand, $p < 0$ is impossible, since induction on D_2 ($q > n$) would yield a decomposition into $D_2 \circ \varphi'' = E'' \# F''$ with $E'' \sim C$, $F'' \sim C^{q-1}$. The minimality of $|p|+|q|$ is then contradicted by the following observations:

(28) (i) $|p+1| + |q-1| = -p + q - 2 < |p|+|q|$;

(ii) $|p+q-1| + 1 = p + q < |q| < |p|+|q|$. □

The following result is an immediate consequence.

(29) Let C be a simple, non-null-homotopic and orientation-preserving closed curve on S. Let $D \sim C^n$, for some $n \geq 2$, with $cr(D) < \infty$. Then either D contains a null-homotopic subcurve or there exists a homeomorphism $\varphi: S_1 \rightarrow S_1$, and there exist closed curves $E \sim C$ and $F \sim C^{n-1}$, so that $D \circ \varphi = E \# F$ and, moreover, so that E is simple.

For orientation-reversing curves C we have a result similar to (27):

(30) Let C be a non-null-homotopic orientation-reversing closed curve on S, and let D be a closed curve, $D \sim C^{2m+1}$, for some $m \geq 1$, with $cr(D) < \infty$. Then there exists a homeomorphism $\varphi: S_1 \rightarrow S_1$ and there exist curves $E \sim C^2$ and $F \sim C^{2m-1}$, so that $D \circ \varphi = E \# F$.

<u>Proof</u>. By induction on $cr(D)$. Let S'' denote the surface generated by C^2. Then S'' is homeomorphic to the annulus. Let D'' denote the lifting of D^2 to S'', then D'' is a <u>closed</u> curve on S'', going around $2m+1$ times. By the previous result we find a homeomorphism φ'' with $D'' \circ \varphi'' = E'' \# F''$ so that $\pi'' \circ E'' \sim C^2$ and $\pi'' \circ F'' \sim C^{4m}$. As $m > 0$ the argument can be repeated to find a homeomorhism Ψ and a split of D'' into $D'' \circ \Psi = E_1 \# F_1 \# E_2 \# F_2$, with subcurves E_1 and E_2 of D'' homotopic to a lifting of C^2 to S''. We may assume without loss of generality that $\Psi^{-1}[\{e^{i\vartheta} | \; 0 \leq \vartheta \leq \frac{1}{2}\pi\}] \subset \{e^{i\vartheta} | \; 0 < \vartheta < \pi\}$. This means that we can split off $\pi'' \circ E_1[S_1]$ from $D[S_1]$, i.e., $D \circ \varphi = E \# F$ with $E \sim C^2$ and $F \sim C^{2m-1}$ for suitably chosen φ, E and F. □

As a convenient tool in deciding whether $cr(C,D) = mincr(C,D)$ for certain closed curves C,D on a compact surface S, we formulate the following lemma:

(31) Let S be a compact surface with universal covering surface S', and let C and D be closed curves on S, so that C is simple and so that each lifting of C to S' and each lifting of D to S' intersect at most once.
Then $cr(C,D) = mincr(C,D)$.

Proof. First for arbitrary closed curves C,D with $cr(C,D) < \infty$ we define an equivalence relation on the points of intersection of C and D. To this end, for any closed curve $D:S_1 \to S$ and $z,z' \in S_1$, let us call a path $P:[0,1] \to S$ a z-z'-*walk along* D if there exist $t,t' \in \mathbb{R}$ so that:

(32) $z = \exp(2\pi it)$, $z' = \exp(2\pi it')$;
$P(x) = D(\exp(2\pi i((1-x)t+xt')))$, for $x \in [0,1]$.

Now let C,D be closed curves on S so that the set $X(C,D) := \{(x,y) \in S_1 \times S_1 \mid C(x) = D(y)\}$ is finite, and so that if $(x,y) \in X(C,D)$ then C and D form crossing (i.e. not touching) curves, if we restrict them to small neighborhoods of x and y. Then the following defines an equivalence relation on $X(C,D)$:

(33) $(x,y) \approx (x',y')$ iff some x-x'-walk along C is homotopic to some y-y'-walk along D.

We call an equivalence class of this relation *odd* if it contains an odd number of elements. Let

(34) $odd(C,D) :=$ number of odd classes of $\approx$.

Now if each lifting of C crosses each lifting of D at most once, then we have that $cr(C,D) = odd(C,D)$. On the other hand we will show below, that $odd(C,D)$ is a lower bound on the number $mincr(C,D)$. This settles our proof, since then $\mathrm{mincr}(C,D) \geq \mathrm{odd}(C,D) = \mathrm{cr}(C,D) \geq \mathrm{mincr}(C,D)$, with equality throughout.

To see that $\mathrm{odd}(C,D)$ indeed bounds $\mathrm{mincr}(C,D)$ from below we use the theory of simplicial approximation. Let $\tilde{C} \sim C$ and $\tilde{D} \sim D$ attain $\mathrm{mincr}(C,D)$. By the tameness of the surface S we may assume that C, $\tilde{C}$, D and $\tilde{D}$ cross each other and themselves a finite number of times. Hence we may assume that C and $\tilde{C}$ follow the edges of a triangulation Γ of S, and that D and $\tilde{D}$ follow the edges of some other triangulation Δ of S, so that Γ and Δ intersect only in edges.

One easily checks that $\mathrm{odd}(C,D)$ is invariant under the following modification of C: if C passes along edges e of triangle T of Γ, replace e by the other two edges of T; similarly for D with respect to Δ. Since $\tilde{C}$ and $\tilde{D}$ arise from C and D by a series of these modifications and their reverses, we have:

(35) $$\mathrm{mincr}(C,D) = \mathrm{cr}(\tilde{C},\tilde{D}) \geq \mathrm{odd}(\tilde{C},\tilde{D}) = \mathrm{odd}(C,D),$$

which proves our claim above. □

Graphs, curves and crossings.

In addition to the previous definitions of the functions cr(.,.) and mincr(.,.) we define, for a path P on S and a closed curve C on S

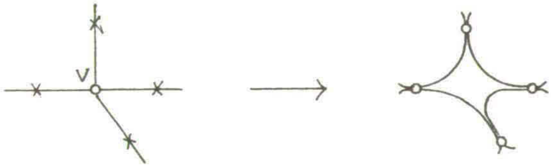
(36) $$\mathrm{cr}(P,C) := |\{(x,z)\in[0,1]\times S_1 \mid P(x)=C(z)\}|,$$

and for a graph G embedded on a surface S, and a closed curve D on S:

(37) $$\mathrm{mincr}(G,D) := \min\{\mathrm{cr}(G,\tilde{D}) \mid \tilde{D} \sim D,\ \tilde{D} \text{ does not intersect } V(G)\}.$$

Blowing up a graph.

An important tool in the study of intersections of curves with a graph G embedded on a surface S is the graph $\overline{G}$ obtained by 'blowing up the vertices in the embedding of G on S until they touch', i.e. until (faces corresponding to) adjacent vertices touch each other. $\overline{G}$ can be defined more precisely as follows. As vertex set of $\overline{G}$ we take $V(\overline{G}) := E(G)$. With each edge $e = \{v,w\}$ of G we identify a point P_e on e (in the embedding on S) so that $P_e \neq v,w$. For each vertex v of G, let $e_0^v, e_1^v, \ldots, e_{d-1}^v$ denote the edges incident with v, in this cyclical order (with respect to the embedding of G on S). Here d denotes the degree of v. Then $P_{e_{i-1}^v}$ and $P_{e_i^v}$ are connected in $\overline{G}$ by an edge the embedding on S of which is homotopic with the path from $P_{e_{i-1}^v}$ to $P_{e_i^v}$ along e_{i-1}^v and e_i^v via v, for i=1,2,...,d (taking indices modulo d):

(38) 

[Note that possibly $P_{e_i} = P_{e_j}$ while $i \neq j$, as loops may occur.]

Clearly $\overline{G}$ is a 4-regular graph, and the faces of $\overline{G}$ with respect to its embedding on S, correspond to vertices and faces of G. For $v \in V(G)$, let disc(v) denote the face of $\overline{G}$ corresponding to v. We can use the term disc as obviously such a face is simply-connected by definition of $\overline{G}$. If f and g are adjacent faces of $\overline{G}$ then one of them corresponds to a vertex of G and the other corresponds to a face of G. [Notice that if each face of G is simply-connected, then blowing up the dual graph G^* of G also yields $\overline{G}$.] It is easy to show that for a closed curve D on S we have

$$\text{(39)} \qquad \operatorname{mincr}(\overline{G},D) = 2\cdot\min\{\, cr(G,\widetilde{D}) \mid \widetilde{D} \sim D \,\}.$$

(So here the minimum ranges over curves $\widetilde{D}$ possibly intersecting V(G).)

SECTION 3 - PROOF OF THE REDUCTION THEOREM.

In this section we finally prove our main theorem.

Reduction Theorem: *Let* S *be a compact surface and let* $G = (V(G), E(G))$ *be a graph embedded on* S. *Let* $C_1, \ldots, C_k$ *be pairwise disjoint closed curves on* S, *each of them simple and orientation preserving.*

If there exists a closed curve D *on* S *satisfying* :

(40) (i) $D \cap V(G) = \emptyset$;

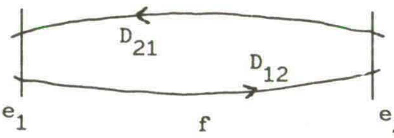
(ii) $cr(G,D) < \Sigma_i \text{mincr}(C_i, D)$,

then there exists one satisfying the further following conditions :

(41) (i) D *is simple* ;

(ii) *the intersection-sequence* $\bar{D}$ *of* D *is simple or semi-simple* ;

(iii) *if* e_1, e_2 *are edges of* G, *and* f *is a face of* G, *so that part* D_{12} *of* D *traverses* f, *going from* e_1 *to* e_2, *and part* D_{21} *traverses* f, *going from* e_2 *to* e_1, *as in* :



then e_1, D_{12}, e_2 *and* D_{21} *enclose a simply connected part of face* f.

Proof of the Reduction Theorem.

Clearly, we may assume that each C_i is non-null-homotopic.

The reduction theorem trivially holds if S is the 2-sphere or the projective plane, since then each orientation preserving closed curve on S is null-homotopic. Then $k = 0$ and no curve satisfies condition (40)(ii). So from now on we assume that:

(42) S is not equal to the 2-sphere or the projective plane.

Now suppose there is a curve satisfying (40). Choose such a curve D, and furthermore choose curves $\tilde{C}_i \sim C_i$, so that each $\tilde{C}_i$ is simple and $cr(\tilde{C}_i,\tilde{C}_j)=0$, for $i \neq j$, so that, in order of priority:

(43) (i) the number of distinct faces and edges traversed by D is minimal;

(ii) $cr(G,D)$ is minimal;

(iii) $\Sigma_i cr(\tilde{C}_i,D)$ is minimal;

(iv) $cr(D)$ is minimal.

We can assume without loss of generality that no point on S is traversed by D and the $\tilde{C}_i$ more than twice. W.l.o.g. $C_i = \tilde{C}_i$.

The remainder of the proof shows that this choice of D (and the $\tilde{C}_i$) guarantees to find a closed curve satisfying the conditions (40) and (41). We will first prove that D is simple. Next we show that D does not cross too many faces too often. Then the order in which D crosses the vertices and faces of G is more or less fixed. Ultimately, we show that if the intersection-sequence $\overline{D}$ of D is not simple, then it overlaps with itself only once. This settles the proof of the Reduction Theorem.

Let Δ denote the set of liftings of D to the universal covering surface S', and let Γ_i denote the set of liftings of C_i to S'. Furthermore, let Γ denote the union of the Γ_i. As D and the C_i are non-null-homotopic and as S is not the projective plane, it follows from (25) that the liftings in Δ and Γ are infinite open curves on S'.

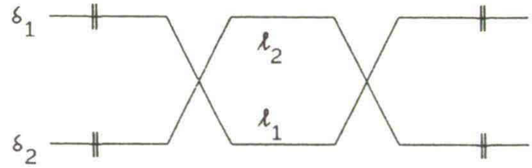
Claim 1. If $\delta \in \Delta$ crosses liftings $\gamma_1,\gamma_2 \in \Gamma$ consecutively, then the intermediate part of δ is simple.

Proof. Suppose $\delta(x) \in \gamma_1[\mathbb{R}]$, $\delta(y) \in \gamma_2[\mathbb{R}]$, for some $x < y$, so that for no ξ with $x < \xi < y$, and for no $\gamma \in \Gamma$ we have $\delta(\xi) \in \gamma[\mathbb{R}]$. Then $y-x \leq 1$, and so if there exist ξ_1,ξ_2 with $x < \xi_1 < \xi_2 < y$, so that $\delta(\xi_1) = \delta(\xi_2)$, then it follows that $\overline{D}$ contains a null-homotopic subcurve (viz. one with $\pi \circ \delta[\xi_1,\xi_2]$ as its image on S). $\square$

<u>Claim 2</u>. Let $\delta_i \in \Delta$ and $\gamma_{i1}, \gamma_{i2} \in \Gamma$ be such that δ_i crosses γ_{i1} and γ_{i2} consecutively, for i=1,2. Let ℓ_i denote the line segment of δ_i between its intersections with γ_{i1} and γ_{i2}.
If $\pi(\ell_1) \neq \pi(\ell_2)$, then ℓ_1 and ℓ_2 cross at most once.

<u>Proof</u>. Suppose that ℓ_1 and ℓ_2 cross at least twice and we have:

(44)

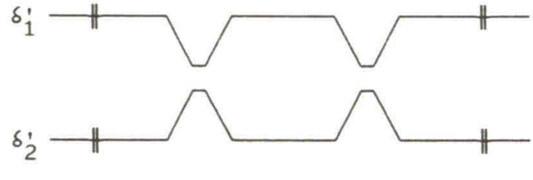


Here $\nparallel$ denotes the intersection of δ_i and γ_{ij}.

Let $x_{ij} \in \mathbb{R}$ be such that $\delta_i(x_{ij}) \in \gamma_{ij}$, i,j = 1,2 and $\delta[x_{i1}, x_{i2}] = \ell_i$, for i = 1,2. Let $u_i \neq v_i$ denote the points of intersection of ℓ_1 and ℓ_2, i.e., let $x_{i1} < u_i < v_i < x_{i2}$, for i = 1,2, so that $\{\delta_1(u_1), \delta_1(v_1)\} = \{\delta_2(u_2), \delta_2(v_2)\}$.

Now, if $\pi \circ \delta_1(x_{11}) \neq \pi \circ \delta_2(x_{21})$, then we can re-route D at points $\pi \circ \delta_1(u_1)$, $\pi \circ \delta_1(v_1)$, $\pi \circ \delta_2(u_2)$ and $\pi \circ \delta_1(v_1)$, so that the intersection of these line segments disappears:

(45)



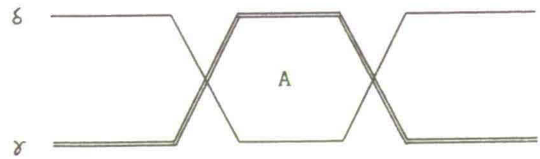
This would yield D' with cr(D') < cr(D), contradicting the minimality of D (and the C_i) as formulated in (43)(iv).
[Remark: If $\pi \circ \delta_1(x_{11}) = \pi \circ \delta_2(x_{21})$, then re-routing is still possible if intervals $[u_1 - x_{11}, v_1 - x_{11}]$ and $[u_2 - x_{21}, v_2 - x_{21}]$ do not overlap.] □

<u>Claim 3</u>. No lifting $\delta \in \Delta$ crosses a lifting $\gamma \in \Gamma$ more than once.

<u>Proof</u>. Suppose there exist liftings $\delta \in \Delta$ and $\gamma \in \Gamma$ crossing more than once. By the disjointness of the liftings in Γ we may assume that δ has two

consecutive intersections with Γ both lying on some $\gamma\in\Gamma$. Let A denote the area enclosed by part of γ and part of δ, cut off by γ:

(46)



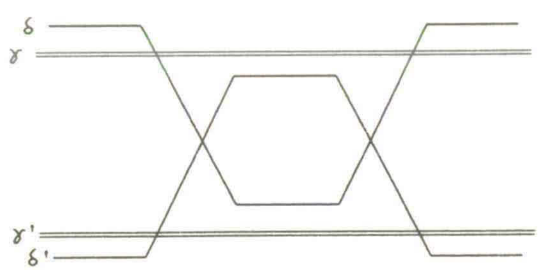
Area A is well defined as by claim 1 δ has no intermediate self-intersections. The liftings in Δ and Γ form the embedding on S' of an infinite graph G'. The area A contains a finite number of faces of G'. We choose δ and γ so that the number of faces enclosed by A is minimum.

Let $\delta(u)=\gamma(x)$ and $\delta(v)=\gamma(y)$ denote the corners of A. As $\delta(u)$ and $\delta(v)$ are consecutive intersections of δ with Γ it follows that $|u-v| \leq 1$. As the C_i are orientation preserving we find that $|u-v| < 1$. Furthermore by the minimality of A, each lifting $\delta'\in\Delta$ entering $\bar{A}$ via γ has to leave $\bar{A}$ via δ. Let $n\in\mathbb{N}$ be so that $n-1 < |x-y| < n$.

If each lifting $\delta'\in\Delta$ entering A via δ leaves A via γ, then re-routing the lifting γ along the δ-boundary of A yields the lifting of a closed curve $E \sim C_i^n$, with $cr(E,\bar{D}) < n\cdot cr(C_i,\bar{D})$. As $cr(E) < \infty$, our lemma (29) applies. After deletion from E of all null-homotopic subcurves we find a simple closed curve $C' \sim C_i$ with $cr(C',\bar{D}) < cr(C_i,D)$, thus contradicting the minimality of D and the C_i formulated in (43)(iii).

Now suppose some lifting $\delta'\in\Delta$ enters and leaves A via $\delta[u,v]$. Then, by the previous claim, there exists a lifting $\gamma'\in\Gamma$ with $\pi\circ\gamma' = \pi\circ\gamma$ so that we have:

(47)



Let $x < u < v$ and $x' < u' < v'$ be such that $\delta(x)\in\gamma$, $\delta'(x')\in\gamma'$ and $\{\delta(u),\delta(v)\} = \{\delta'(u'),\delta'(v')\}$. Then w.l.o.g. $x=x'$ and $(u,v)\cap(u',v') \neq \emptyset$

(cf. the remark at the end of the proof of the previous claim). Let A' denote the area enclosed by δ' and γ', then $\pi \circ A = \pi \circ A'$. Using claim 1 once again, we may assume that $\delta[u,v]$ and $\delta'[u',v']$ intersect only at their common endpoints and that they enclose an area $A'' \subset A \cap A'$. This enables us to use the classical fixed-point theorem of Brouwer [1]. For instance, let $\Psi:\mathbb{C}\rightarrow\mathbb{C}$ denote the 1-1 transformation of the universal covering surface S' onto itself with $\Psi[A] = A'$. By the 'uniqueness' of the universal covering surface Ψ has no fixed point. Furthermore let $\Phi:A'\rightarrow A''$ be a continuous map with $\Phi(x)=x$, for $x \in A''$, and $\Phi(x) \in \delta[u,v]$, for $x \notin A''$. Then the composition $\Phi \circ \Psi : A \rightarrow A''$ should have a fixed point, by Brouwer's theorem. This may lead to a contradiction.

Applying this technique we find first that the intervals (u,v) and (u',v') are not nested. A second application shows that $\delta(u)=\delta'(u')$ and $\delta(v)=\delta'(v')$, and that we may assume that $u' < u < v' < v$. Assuming that, under minimality of A and of A'', u was chosen to be minimal, we finally arrive at a contradiction, showing that (47) does not occur. This settles the proof of our claim. □

Remark. This intermediate result is quite useful. It follows from lemma (31) that $cr(C_i,D) = mincr(C_i,D)$, for all i. In the remainder of the proof, this fact is exploited to 'break down' D as much as possible.

Claim 4. Each $\delta \in \Delta$ is simple.

Proof. Suppose $\delta \in \Delta$ and $x < y$ are such that $\delta(x) = \delta(y)$. A lifting $\gamma \in \Gamma$ crossing $\delta[x,y]$ has to intersect δ at least twice. As this is impossible by the previous claim, we have that $y-x < 1$ and we can split off a null-homotopic subcurve (with image $\pi \circ \delta[x,y]$ on S) from D, thus contradicting the minimality of D w.r.t. condition (43)(iv). □

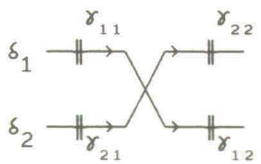
Claim 5. D is simple.

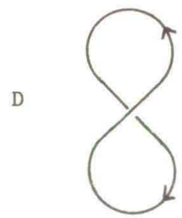
Proof. By the previous claims we know that any lifting δ of D to S' passes from one lifting $\gamma \in \Gamma$ to another. By the disjointness of the C_i this means that each $\delta \in \Delta$ crosses each $\gamma \in \Gamma$ at most once. Hence $cr(C_i,D) = mincr(C_i,D)$ for all i (by lemma (31)). If D has a self-intersection in point P, there are two ways of re-routing D in this point P. Working out an idea of Lovász and Seymour we show that at least one of these ways of uncrossing D in point P yields again a configuration with a minimum number of intersections with the C_i. Actually, we show in the following, that a crossing of distinct liftings δ and δ' of D to S' can be removed under preservation of the property of 'jumping' from one lifting $\gamma \in \Gamma$ to another.

Suppose liftings δ_1, δ_2 of D cross in a point $\delta_1(x) = \delta_2(y) = p$, say. Without loss of generality $D = D_1 \# D_2$, with $D_1(1)=\pi(p)$, $D_1[S_1] = \pi \circ \delta_1[x,x+\frac{1}{2}]$ and $D_2[S_1] = \pi \circ \delta_2[y,y+\frac{1}{2}]$. Let γ_{i1}, γ_{i2} denote liftings in Γ crossed by δ_i just before, respectively just after traversing point p. W.l.o.g., $\Sigma_j cr(C_j, D_1) > 0$. We distinguish two cases.

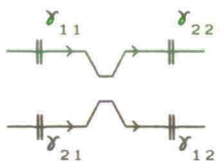
<u>Case 1</u>: $\Sigma_j cr(C_j, D_i) > 0$, for i=1,2.

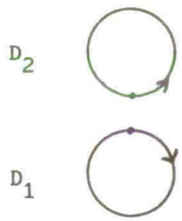
Let $x_{ij} \in \mathbb{R}$ be so that $\delta_i(x_{ij}) \in \gamma_{ij}$, for i=1,2, j=1,2. Then obviously $\pi \circ \delta_1[x_{11}, x_{12}]$ and $\pi \circ \delta_2[x_{21}, x_{22}]$ are distinct subpaths of D. If $\gamma_{11} = \gamma_{21}$ or $\gamma_{12} = \gamma_{22}$, then $\gamma_{11} \neq \gamma_{22}$ and $\gamma_{12} \neq \gamma_{21}$, so the liftings of the curves D_1 and D_2, arising when

(48) 

corresponding to: 

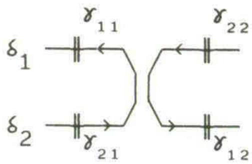
is replaced by :

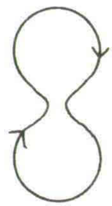
(49) 

corresponding to: 

again pass from one lifting in Γ to another. Hence at least one of them yields an example smaller than D in the sense of (43)(i).

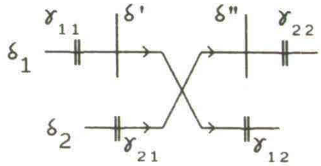
If on the other hand $\gamma_{11} \neq \gamma_{21}$ and $\gamma_{12} \neq \gamma_{22}$, then changing D, so as to obtain a curve with liftings passing from γ_{21} to γ_{11} or from γ_{22} to γ_{12} as in:

(50) 

corresponding to: $D_1 \# (D_2)^{-1}$ 

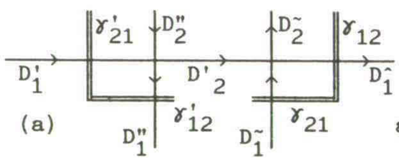
yields a curve $D_1 \# (D_2)^{-1}$ that is a smaller example in the sense of (43)(iv).

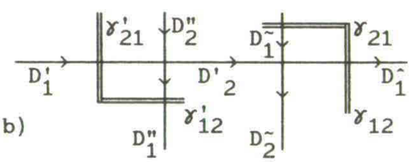
<u>Case 2</u>: $\Sigma_j cr(C_j, D_1) > 0 = \Sigma_j cr(C_j, D_2)$.
We are in the following situation:

(51) 

If $\gamma_{21} \neq \gamma_{12}$, then D_1 provides an example smaller than D.

If $\gamma_{21} = \gamma_{12}$, then $D_1 \# (D_2)^{-1}$ is easily seen to contradict the minimality of D, as in both cases:

(52) (a) 

and (b) 

the lifting of $D_1 \# (D_2)^{-1}$ passing along D'_2 'jumps' from one lifting $\gamma \in \Gamma$ to another (since $\gamma'_{12} = \gamma'_{11} \neq \gamma_{12} = \gamma_{21}$).

The reasoning above shows that distinct liftings of D cannot cross. By the previous claim a lifting cannot cross itself. This settles our claim that cr(D) = 0. □

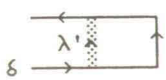
Next we show that D does not traverse a face of G too often. We study a face f of G that is traversed by D more than once, and we find restrictions on the way the curve D may traverse such a face.

For convenience, we first introduce a short-hand notation for a frequently appearing sum of crossing numbers. For a fixed choice of the curves C_i, and for any path, curve or graph X, embedded on S the C-crossing number CC(.) is defined by:

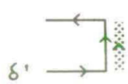
(53) $$CC(X) := \Sigma_i cr(X,C_i) .$$

<u>Claim 6</u>. Let λ be a simple path inside a face f of G, with end points on distinct components of $D[S_1]\cap f$, and such that it has no other intersections with D.
Then each lifting λ' of λ to S' has its end points on distinct liftings of D.

<u>Proof</u>. Let λ' denote a lifting of λ to S', intersecting liftings $\delta \in D$ at $\lambda'(0)$ and $\lambda'(1)$. As δ does not cross itself we have:

(54) 

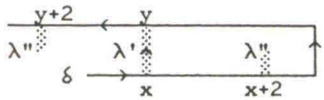
Let $x < y$ be such that $\delta(x)=\lambda'(0)$ and $\delta(y)=\lambda'(1)$. Notice that $\pi\circ\delta(x) = \pi\circ\lambda'(0) \neq \pi\circ\lambda'(1) = \pi\circ\delta(y)$. So if $y-x \leq 1$, then $y-x < 1$ and D can be short-cut, by passing along λ:

(55) 

thus contradicting the minimality of D w.r.t. conditions (43)(i)-(ii).

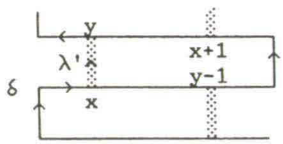
If $y-x \geq 2$, then $y-x > 2$ and we have a lifting λ'' of λ, with $\lambda''(0) = \delta(x+2)$. This leads to a contradiction as distinct liftings of λ are disjoint, as liftings of D do not cross themselves and as a lifting of D does not cross the interior of any lifting of λ:

(56)



These arguments apply also to x+1 in case D is orientation-preserving. On the other hand, if D is orientation-reversing, and we have that $x < y-1 < x+1 < y$, then the following situation may occur:

(57)

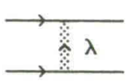


Without loss of generality, D decomposes into $\text{path}(D) = P\#Q$, with $P[0,1] = \pi\circ\delta[x,y-1]$ and $Q[0,1] = \pi\circ\delta[y-1,x+1]$. Under the assumption that the C_i are chosen so that $CC(\lambda)$ is minimal, it follows that the curve D', defined by $\text{path}(D') = \lambda\#Q$ has liftings to S' 'jumping' from one $\gamma\in\Gamma$ to another, hence D' is an example contradicting the minimality of D. □

<u>Claim 7</u>. Let λ be a simple path inside a face f of G, with end points on $D[S_1]$, and such that it has no other intersections with D. Then with respect to the orientation of λ, D intersects λ once from left to right and once from right to left.

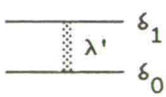
<u>Proof</u>. Suppose D hits λ twice from the left:

(58)



and assume that the C_i are chosen in such a way that (43) holds, and that under this condition $CC(\lambda)$ is minimal. Consider a lifting λ' of λ:

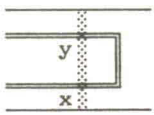
(59)



Here δ_0,δ_1 denote distinct liftings of D intersecting λ' in λ'(0) and λ'(1), respectively. As λ and the C_i are simple, and as no lifting $\delta\in D$

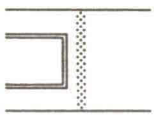
crosses a lifting $\gamma \in \Gamma$ more than once, it follows from the minimality of $CC(\lambda)$ that no lifting $\gamma' \in \Gamma$ crosses λ' more than once. To see this, suppose to the contrary there is a curve C_i, the lifting γ' of which crosses λ' at least twice. Consider a minimal example, i.e., one with λ' and γ' enclosing a minimum number of faces of the infinite graph obtained by lifting λ and the C_i to S'. Let $\gamma'(x)=\lambda'(u)$ and $\gamma'(y)=\lambda'(v)$ denote the intersection points of λ' and γ'. Without loss of generality we have $x < y$ and the situation is as follows:

(60)



Obviously $y-x < 1$, so the curve can be re-routed decreasing $CC(\lambda)$:

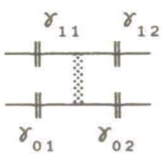
(61)



a contradiction.

Let γ_{i1} and γ_{i2} denote the liftings in Γ crossing δ_i just before and just after traversing the endpoint $\lambda'(i)$, for i=0,1:

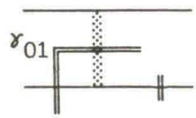
(62)



Without loss of generality, let D be decomposed into path(D) = $P_0 \# P_1$, with $P_i(0) = \lambda(i)$, for i=0,1.

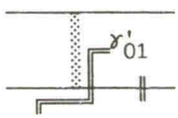
If for both i=0,1 we have $CC(P_i) > 0$, then we know by the minimality of $CC(\lambda)$ that none of the liftings γ_{ij}, i=1,0, j=1,2, cross the lifted path λ'. To see this, suppose that γ_{01} crosses λ'. So we have:

(63)



In this case we can re-route the simple closed curve with image $\pi \circ \gamma_{01}[\mathbb{R}]$ so that the resulting curve is again simple and disjoint from the other C_i, while it has the same number of intersections with D and so that $CC(\lambda)$ decreases by one:

(64)



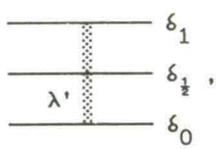
Therefore none of $\gamma_{11}, \gamma_{12}, \gamma_{01}, \gamma_{02}$ intersect λ'. It follows that at least one of the curves D_i, defined by $\text{path}(D_1) := \lambda \# P_1$, $\text{path}(D_0) := \lambda^{-1} \# P_0$, yields an example smaller than D.

If we have $CC(P_1) = 0$ (hence $CC(P_0) > 0$), then by arguments similar to the above, none of γ_{11} and γ_{02} cross λ'. Now it follows that D_0 defined above is an example smaller than D. $\square$

Claim 8. Let λ denote a simple path inside a face f of G, intersecting D only in $\lambda(0)$, $\lambda(\frac{1}{2})$ and $\lambda(1)$ at distinct components ℓ_1, ℓ_2, ℓ_3 of $D \cap f$. Then ℓ_2 does not separate $\lambda[0,\frac{1}{2}]$ from $\lambda[\frac{1}{2},1]$.

Proof. Let λ be decomposed into $\lambda = \lambda_1 \# \lambda_2$. Assume that the C_i are chosen in such a way that (43) holds, and that under this condition $CC(\lambda)$ is minimum. Consider liftings in D and Γ and some lifting λ' of λ:

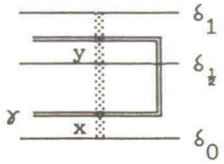
(65)



where δ_i denotes the lifting of D intersecting λ' in $\lambda'(i)$, for $i = 0, \frac{1}{2}, 1$. It follows from the previous claims, that these liftings of D are distinct and that they hit λ' alternatingly from the left and from the right.

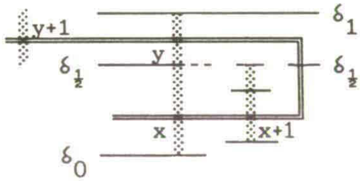
It is easily verified, using the same arguments as in the proof of the previous claim, that there is no lifting $\gamma \in \Gamma$ crossing $\lambda'[0,\frac{1}{2}]$ or $\lambda'[\frac{1}{2},1]$ more than once. Now suppose that we have:

(66)



If $y-x < 1$ we can re-route γ (and hence the corresponding C_i) along λ', thus decreasing $CC(\lambda)$. If $y-x > 1$, we consider a lifting λ'' of λ with $\lambda''(0) = \gamma'(x+1)$. As the C_i are orientation-preserving we find ourselves in the following situation:

(67)



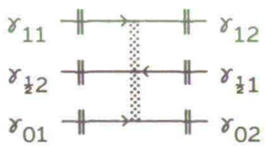
.

So λ'' should intersect λ', a contradiction. Hence (66) does not occur, and therefore we know that

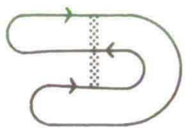
(68) no lifting $\gamma \in \Gamma$ intersects λ' more than once.

We may assume without loss of generality, that the liftings of D are directed as in

(69)



corresponding to

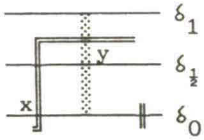


Here γ_{i1} and γ_{i2} denote the liftings in Γ crossed by the lifting δ_i of D just before and just after traversing $\lambda'(i)$, for $i=0,\frac{1}{2},1$.

Without loss of generality let D be decomposed into path(D) $= P_0 \# P_{\frac{1}{2}} \# P_1$, with $P_i(0) = \lambda(i)$, for $i = 0,\frac{1}{2},1$. We now show:

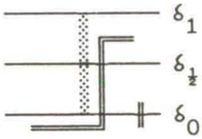
(70) (i) If $CC(P_1) > 0$, then γ_{01} does not intersect λ';
(ii) If $CC(P_0) > 0$, then γ_{02} does not intersect λ';
(iii) If $CC(P_{\frac{1}{2}}) > 0$, then γ_{11} does not intersect λ';
(iv) If $CC(P_1) > 0$, then γ_{12} does not intersect λ'.

By symmetry it suffices to prove (70)(i). By the arguments used in the proof of the previous claim, it follows easily that if the simple closed curve with lifting γ_{01} crosses the path P_1, then γ_{01} cannot cross $\lambda'[0,\frac{1}{2}]$. Furthermore, γ_{01} cannot cross $\lambda'[\frac{1}{2},1]$. For suppose we have $\gamma_{01}(x) \in \delta_0$ and $\gamma_{01}(y) \in \lambda'[\frac{1}{2},1]$, and we are in the following situation:

(71) 

.

Then it is easy to see that $y-x < 1$, so the curve with lifting γ_{01} can be re-routed along λ, thus decreasing $CC(\lambda)$:

(72) 

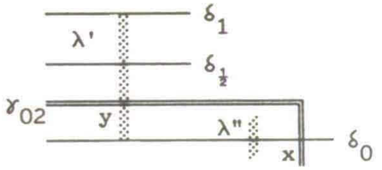
Analogous results hold for $\pi\circ\gamma_{02}$ crossing P_0 ; $\pi\circ\gamma_{11}$ crossing $P_{\frac{1}{2}}$; and $\pi\circ\gamma_{12}$ crossing P_1. This shows (70).

We next show:

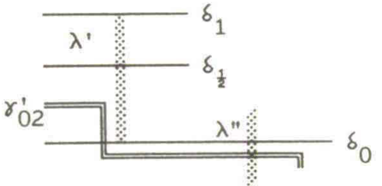
(73) (i) If $CC(P_0) = 0$ and $CC(P_{\frac{1}{2}}) > 0$, then γ_{02} does not intersect λ';
(ii) If $CC(P_{\frac{1}{2}}) = 0$ and $CC(P_0) > 0$, then γ_{11} does not intersect λ'.

By symmetry it suffices to show (73)(i). Suppose that we have that $\pi(\gamma_{02}\cap\delta_0)$ is on $P_{\frac{1}{2}}$. Then first we claim that γ_{02} does not cross $\lambda'[0,\frac{1}{2}]$.

To see this, suppose that we have $x<y$, with $\gamma_{02}(x) \in \delta_0$ and $\gamma_{02}(y) \in \lambda'[0,\frac{1}{2}]$, and we are in situation

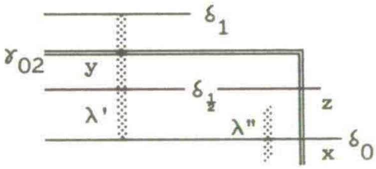
(74) 

Here λ'' denotes a lifting of λ, with $\lambda''(\frac{1}{2}) \in \delta_0$. It is obvious that $\gamma_{02}[x,y]$ crosses λ'' and that $y-x < 1$, so the curve with lifting γ_{02} can be re-routed along D :

(75) 

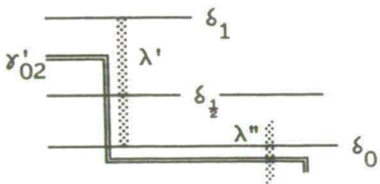
hereby decreasing $CC(\lambda)$. So γ_{02} does not cross $\lambda'[0,\frac{1}{2}]$.

Second, we claim that γ_{02} does not cross $\lambda'[\frac{1}{2},1]$. For suppose we have:

(76) 

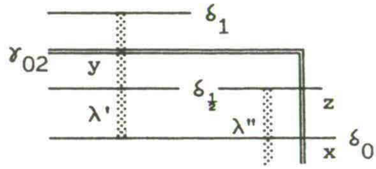
Here $\gamma_{02}(z)$ denotes the intersection of γ_{02} and $\delta_{\frac{1}{2}}$. If $y-x > 1$, then it follows that $z = x+1$ (since $\gamma_{02}(x,y)$ can intersect no other $\delta \in D$ than $\delta_{\frac{1}{2}}$, and since γ_{02} has an intersection with some $\delta \in D$ at $z=x+1$). This, however, is impossible, as δ_0 and $\delta_{\frac{1}{2}}$ go in opposite directions, and the C_i are orientation-preserving. Hence $y-x < 1$.

If λ'' crosses $\gamma_{02}[x,y]$, then the curve with lifting γ_{02} can be re-routed as in:

(77) 

contradicting the minimality of $CC(\lambda)$.

If λ'' does not cross $\gamma_{02}[x,y]$, then one of the end points of λ'' is on $\delta_{\frac{1}{2}}$:

(78) 

It follows that $\lambda''\cap\delta_{\frac{1}{2}} = \{\lambda''(0)\}$. Hence the closed curve D' on S defined by $path(D') = P_0 \# \lambda_1^{-1}$ has the property that $(D')^2$ is null-homotopic. As S is not the projective plane we know, by (25), that D' itself is null-homotopic. This however contradicts the minimality of D, since it means that D may be short-cut by replacing part P_0 of D by λ_1. So γ_{02} does not cross $\lambda'[\frac{1}{2},1]$. This finishes the proof of (73).

Taking together the intermediate results (68), (70) and (73), we get:

(79) (i) no lifting $\gamma\in\Gamma$ intersects λ' more than once;

(ii) if $CC(P_0 \# P_{\frac{1}{2}}) > 0$, then γ_{02} and γ_{11} do not cross λ';

(iii) if $CC(P_1) > 0$, then γ_{01} and γ_{12} do not cross λ'.

It follows that at least one of the curves D_0, D_1 defined by

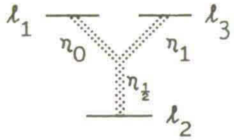
(80) $$path(D_0) = P_0 \# P_{\frac{1}{2}} \# \lambda^{-1}; \; path(D_1) = P_1 \# \lambda ,$$

yields an example contradicting the minimality of D, as formulated in (43)(i)-(ii). □

The previous claim asserts that distinct components ℓ_1, ℓ_2, ℓ_3 of $D\cap f$ cannot be lined up one after the other. However, the ℓ_i may have other relative positions. The next claim deals with this case.

<u>Claim 9</u>. Each face f of G is traversed by D at most twice.

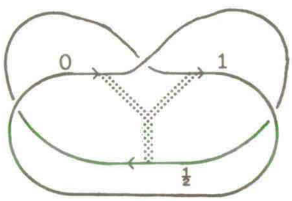
Proof. Suppose f is a face of G so that $D \cap f$ has more than two components. Let λ denote a simple path inside face f, intersecting D only in $\lambda(0)$, $\lambda(\frac{1}{2})$ and $\lambda(1)$ at distinct components ℓ_1, ℓ_2, ℓ_3 of $D \cap f$. By the previous claim we know that ℓ_2 does not separate $\lambda[0,\frac{1}{2}]$ from $\lambda[\frac{1}{2},1]$. It is obvious that we can construct a Y-shaped connection Y between the points $\lambda(i)$. That is, there exists a point p on S, and there exist simple and pairwise internally disjoint paths η_i on S with $\eta_i(0) = p$ and $\eta_i(1) = \lambda(i)$, for $i = 0,\frac{1}{2},1$. We have, without loss of generality:

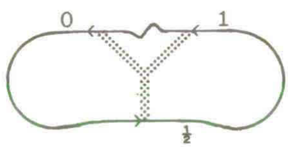
(81) Y : 

We consider liftings Y', η_0', $\eta_{\frac{1}{2}}'$, η_1' of Y, η_0, $\eta_{\frac{1}{2}}$ and η_1 respectively. Assume that the C_i are chosen so that under the minimality condition (43) Y has a minimum number of crossings with the C_i (CC(Y) is minimum). Without loss of generality, let D be decomposed into:

(82) $\quad \text{path}(D) = P_0 \# P_{\frac{1}{2}} \# P_1$

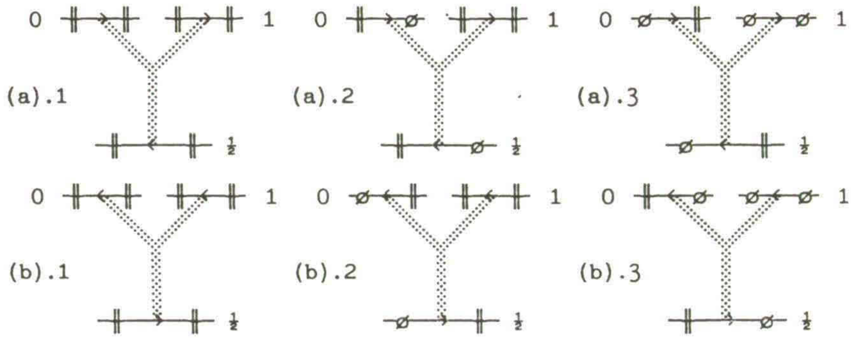
with $P_i(0) = \eta_i(1)$, for $i = 0,\frac{1}{2},1$. Let $\delta_i \in D$ denote the lifting of D crossing Y' at $\eta_i'(1)$. By claim 6 the δ_i are distinct and, by claim 7, they have the same orientation seen from the point of view of point p. We have:

(83) (a): 

or (b): 

Let $\gamma_{i1}, \gamma_{i2} \in \Gamma$ denote the liftings crossed by δ_i just before and just after traversing $\eta_i'(1)$. We have to distinguish several cases taking into account whether $CC(P_i)$ is equal to zero or not, for $i = 0,\frac{1}{2},1$. By symmetry we can restrict ourselves to 6 cases:

(84)

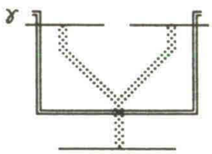


In this scheme the lifting P' of a path P_i has been marked with ǂ if $CC(P_i) > 0$, and with $\varnothing$ otherwise.

If there are ties, case (b).2 is preferred to case (b).3. By minimality of CC(Y) we know that there is no lifting $\gamma \in \Gamma$ crossing Y' more than once. Furthermore no lifting γ_{ij} marked with ǂ in our scheme crosses η_i' (as is proved using the arguments of claim 7).

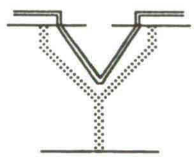
Suppose the following occurs:

(85)



where ǂ has the same meaning as in the scheme above. Then the number CC(Y) is decreased by re-routing the closed curve with lifting γ along the other 'legs' of Y:

(86)



This shows:

(87) situation (85) does not occur.

We now consider the six cases (84).

Case (a).1. $CC(P_i) \neq 0$, for $i = 0,\frac{1}{2},1$. By (87) we have that $\gamma_{01} \neq \gamma_{12}$. Furthermore γ_{01} does not cross η_0' or η_1' and the same holds for γ_{12}. As there is no lifting $\gamma \in \Gamma$ crossing the path $(\eta_0')^{-1} \# \eta_1'$ twice, it is clear that passing along $(\eta_0')^{-1} \# \eta_1'$, going from γ_{01} to γ_{12} no lifting in Γ is crossed twice. So the lifting of the curve D_0 given by $\mathrm{path}(D_0) = \eta_0^{-1} \# \eta_1 \# P_1$ passes from one $\gamma \in \Gamma$ to another, hence $\Sigma_i \mathrm{mincr}(C_i,D_0) = \Sigma_i \mathrm{cr}(C_i,D_0)$. Analogous results hold for the pair $\gamma_{\frac{1}{2}1},\gamma_{02}$ and for $\gamma_{11},\gamma_{\frac{1}{2}2}$. It follows directly that one of the curves D_i defined by:

(88)
$$\mathrm{path}(D_0) = \eta_0^{-1} \# \eta_1 \# P_1 \ ;$$
$$\mathrm{path}(D_{\frac{1}{2}}) = \eta_{\frac{1}{2}}^{-1} \# \eta_0 \# P_0 \ ;$$
$$\mathrm{path}(D_1) = \eta_1^{-1} \# \eta_{\frac{1}{2}} \# P_{\frac{1}{2}}$$

is an example contradicting the minimality of D, as $\Sigma_i \Sigma_j \mathrm{mincr}(C_j,D_i) = \Sigma_i \Sigma_j \mathrm{cr}(C_j,D_i) \geq \Sigma_j \mathrm{cr}(C_j,D) = \Sigma_j \mathrm{mincr}(C_j,D) > \mathrm{cr}(G,D) = \Sigma_i \mathrm{cr}(G,D_i)$.

Case (a).2. $CC(P_i) = 0$, only for $i=0$. We find that one of the curves D_0,D_1 defined in (88) is smaller than D.

Case (a).3. $CC(P_{\frac{1}{2}}) = CC(P_1) = 0$. We find that $D_{\frac{1}{2}}$ defined in (88) is smaller than D.

Case (b).1. $CC(P_i) \neq 0$, for $i = 0,\frac{1}{2},1$. In analogy to case (a).1 we find that γ_{01} and $\gamma_{\frac{1}{2}2}$ are distinct and do not cross $(\eta_0')^{-1} \# \eta_{\frac{1}{2}}'$, with similar results for $\gamma_{\frac{1}{2}1},\gamma_{12}$ and for γ_{11},γ_{02}. It follows that one of the curves D_i', defined by:

(89)
$$\mathrm{path}(D_0') = \eta_0^{-1} \# \eta_{\frac{1}{2}} \# P_{\frac{1}{2}} \# P_1 \ ;$$
$$\mathrm{path}(D_{\frac{1}{2}}') = \eta_{\frac{1}{2}}^{-1} \# \eta_1 \# P_1 \# P_0 \ ;$$
$$\mathrm{path}(D_1') = \eta_1^{-1} \# \eta_0 \# P_0 \# P_{\frac{1}{2}} \ ,$$

forms an example smaller than D, as we have $\Sigma_i\Sigma_j \mathrm{mincr}(C_j,D_i') = \Sigma_i\Sigma_j \mathrm{cr}(C_j,D_i') \geq 2\cdot\Sigma_j \mathrm{cr}(C_j,D) = 2\cdot\Sigma_j \mathrm{mincr}(C_j,D) > 2\cdot \mathrm{cr}(G,D) = \Sigma_i \mathrm{cr}(G,D_i')$.

Case (b).2. $CC(P_i) = 0$, only for $i=0$. Again $\gamma_{01} \neq \gamma_{\frac{1}{2}2}$ and neither of these liftings cross the path $(\eta_0')^{-1}\#\eta_{\frac{1}{2}}$. Hence D_0' defined in (89) is an example smaller than D.

Case (b).3. $CC(P_{\frac{1}{2}}) = CC(P_1) = 0$. We find that $CC(P_0) = \Sigma_j \mathrm{cr}(D,C_j) > \mathrm{cr}(G,D) \geq 3$. By preference of (b).2 to (b).3 it follows that liftings γ_{02} and $\gamma_{\frac{1}{2}1}$ do not cross η_0' or $\eta_{\frac{1}{2}}'$ (since if, say, γ_{02} would cross η_1', then shifting γ_{02} 'over' η_0' would create an instance of case (b).2). Hence, if $\gamma_{02} \neq \gamma_{\frac{1}{2}1}$, then the curve D" defined by:

$$(90) \qquad \mathrm{path}(D") = P_0\#\eta_{\frac{1}{2}}^{-1}\#\eta_0$$

is an example smaller than D.

So we may assume $\gamma_{02} = \gamma_{\frac{1}{2}1}$. We now claim that one of $D_{\frac{1}{2}}'$ and D_1', defined in (89) yields an example smaller than D. For suppose that both $D_{\frac{1}{2}}'$ and D_1' have liftings crossing some lifting in Γ more than once. We will now exploit the fact that none of the paths $\eta_1^{-1}\#\eta_{\frac{1}{2}}\#P_{\frac{1}{2}}$, $\eta_0^{-1}\#\eta_1\#P_1$ and $\eta_0^{-1}\#\eta_{\frac{1}{2}}\#P_{\frac{1}{2}}\#P_1$ are null-homotopic (by minimality of D).

If $\gamma = \gamma_{02} = \gamma_{\frac{1}{2}1}$ is a lifting of say C_1, then the part of γ between its intersections with P_1 corresponds with part C_1' of C_1. Let $C_1'' := C_1\backslash C_1'$. As any lifting of $D_{\frac{1}{2}}'$ crosses some lifting in Γ more than once, we know that $\eta_{\frac{1}{2}}^{-1}\#\eta_1\#P_1$ is homotopic with C_1' or C_1''. (Using the simplicity of $D_{\frac{1}{2}}'$.) Similarly we find that $P_{\frac{1}{2}}\#\eta_1^{-1}\#\eta_0$ is homotopic with C_1' or C_1''. As $\eta_0^{-1}\#\eta_{\frac{1}{2}}\#P_{\frac{1}{2}}\#P_1$ is non-null-homotopic, we may assume $\eta_{\frac{1}{2}}^{-1}\#\eta_1\#P_1 \sim C_1'$. But then it follows that $\eta_0^{-1}\#\eta_1\#P_1$ is null-homotopic, a contradiction. This settles the last of the six cases, thereby concluding the proof of our claim. $\square$

Next we study the order in which D may traverse faces of G for a second time. It turns out that this order is quite restricted.

Claim 10. If D traverses distinct faces f,g of G twice, then the intersection-sequence $\overline{D}$ of D cannot contain f and g in alternating (cyclic) order, i.e., $\overline{D}$ does not look like: $(\cdots,f,\cdots,g,\cdots,f,\cdots,g,\cdots)$.

Proof. Suppose D traverses faces f and g in alternating cyclic order. Let ℓ_1 and ℓ_3 denote the components of $D\cap f$ and let ℓ_2 and ℓ_4 denote the components of $D\cap g$, so that D contains sub-paths $\ell_1,\ell_2,\ell_3,\ell_4$ (in this cyclic order). Let λ denote a simple path inside face f with endpoints on ℓ_1 and ℓ_3, and with no other intersections with D. Let μ denote the analogous path inside face g. We have without loss of generality:

(91)

Let D decompose into $\text{path}(D) = P_1 \# P_2 \# P_3 \# P_4$, with $P_1(0)=\lambda(0)$, $P_2(0)=\mu(0)$, $P_3(0)=\lambda(1)$, $P_4(0)=\mu(1)$. Assume that the C_i are chosen so that under condition (43) the number of intersections of λ and μ with the C_i is minimal. We prove that at least one of the curves D_1,D_2 defined by

$$\text{(92)} \qquad \text{path}(D_1) = P_1\#\mu\#P_3^{-1}\#\lambda^{-1}; \quad \text{path}(D_2) = P_2^{-1}\#\mu\#P_4\#\lambda \ ,$$

yields an example contradicting the minimality of D. To see this, we will show below:

$$\text{(93)} \qquad \begin{aligned} &\text{(i) } \Sigma_j \text{mincr}(P_1\#\mu\#P_3^{-1}\#\lambda^{-1},C_j) \ \geq \ \Sigma_j(\text{cr}(P_1,C_j)+\text{cr}(P_3,C_j)) \ ; \\ &\text{(ii) } \Sigma_j \text{mincr}(P_2^{-1}\#\mu\#P_4\#\lambda \ ,C_j) \ \geq \ \Sigma_j(\text{cr}(P_2,C_j)+\text{cr}(P_4,C_j)) \ . \end{aligned}$$

Having shown this, it follows that

$$\text{(94)} \qquad \begin{aligned} &\Sigma_j \text{mincr}(D_1,C_j) + \Sigma_j \text{mincr}(D_2,C_j) = \\ &\Sigma_j \text{mincr}(P_1\#\mu\#P_3^{-1}\#\lambda^{-1},C_j) + \Sigma_j \text{mincr}(P_2^{-1}\#\mu\#P_4\#\lambda \ ,C_j) \geq \\ &\Sigma_j(\text{cr}(P_1,C_j)+\text{cr}(P_3,C_j)) + \Sigma_j(\text{cr}(P_2,C_j)+\text{cr}(P_4,C_j)) = \Sigma_j \text{cr}(D,C_j) \geq \\ &\Sigma_j \text{mincr}(D,C_j) > \text{cr}(G,D) = \text{cr}(G,D_1) + \text{cr}(G,D_2). \end{aligned}$$

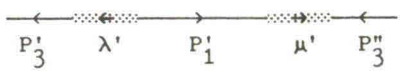
This proves our claim, since D_1 and D_2 are obviously 'smaller' than D.

We now show (93). By symmetry it suffices to prove (93)(i).

In case both $CC(P_1)$ and $CC(P_3)$ are positive, we can use the arguments of the previous claims to show that each lifting of D_1 to S' passes from one lifting $\gamma \in \Gamma$ to another, which settles the proof.

Assuming that $0 = CC(P_1) < CC(P_3)$, we consider a lifting P_1' of P_1 to S', and liftings λ' and μ' of λ and μ, with $\lambda'(0)=P_1'(0)$ and $\mu'(0)=P_1'(1)$. Let P_3' and P_3'' denote liftings of P_3 so that we have :

(95)

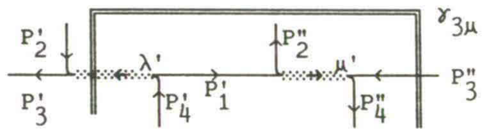


Let $\gamma_{3\lambda}$ denote the lifting in Γ crossing P_3' in $P_3'(u)$ with u minimal, and let $\gamma_{3\mu} \in \Gamma$ cross P_3'' in $P_3''(v)$ with v maximal. Using the facts that the curves C_j are simple and disjoint, that D is simple, and that no pair of liftings $\delta \in D$, $\gamma \in \Gamma$ has more than one intersection, it easily follows that $\gamma_{3\lambda} \neq \gamma_{3\mu}$. We leave this to the reader. We show:

(96) $\gamma_{3\mu}$ does not cross λ'.

To see this, suppose to the contrary that $\gamma_{3\mu}$ does cross λ', so we have:

(97)



Here P_i' and P_i'' denote liftings of P_i, for i=1,2,3,4. It is easily verified given the position of $\gamma_{3\mu}$ and the fact that $\gamma_{3\mu}$ does not intersect the lifting in D containing P_1' more than once, that P_2', P_2'', P_4' and P_4'' cannot be positioned otherwise. Let x,y be such that $\gamma_{3\mu}(x) \in P_3''$ and $\gamma_{3\mu}(y) \in \lambda'$. Without loss of generality $x < y$. It is clear that part $\gamma_{3\mu}[(x,y)]$ of $\gamma_{3\mu}$ is crossed by a lifting of D only once, say in $\gamma_{3\mu}(z)$. Hence $y-x < 2$.

By minimality of $CC(\lambda \cup \mu)$ we know that no lifting $\gamma \in \Gamma$ crosses both λ' and μ'. So if $y-x < 1$, then the curve with lifting $\gamma_{3\mu}$ can be re-routed along part of D_1, not increasing the number of intersections $\Sigma_j cr(D,C_j)$, while decreasing $CC(\lambda \cup \mu)$, a contradiction. On the other hand the assumption that $1 \leq y-x \leq 2$ leads to $z = x+1$, which is contradicted by the fact that the C_j are orientation-preserving. This proves (96).

For reasons of symmetry it follows directly that we also have that $\gamma_{3\lambda}$ does not cross μ'. So we have that for each lifting δ of D_1, and for each $x \in \mathbb{R}$, the number of liftings $\gamma \in \Gamma$ that cross δ once, intersecting δ on $\delta([x,x+1))$, is not less than $\Sigma_j cr(P_3,C_j)$. □

Claim 11. Let D traverse distinct faces f and g of G twice, and let P and Q denote distinct subpaths of D with intersection-sequences $\overline{P}$ and $\overline{Q}$, respectively, so that the intersection-sequence $\overline{D}$ of D contains $\overline{P}$ and $\overline{Q}$ as subsequences:

$$\overline{D} = (\cdots,\underbrace{f,\cdots,g}_{\overline{P}},\cdots,\underbrace{g,\cdots,f}_{\overline{Q}},\cdots).$$

Then $\overline{P} = (\overline{Q})^{-1}$.

Proof. Without loss of generality, let D decompose into path(D)= $P\#A\#Q\#B$, where A is a simple path with both end points in face f, and B is a simple path with both end points in face g. Let λ (μ) denotes a simple path from A(1) to A(0) (from B(1) to B(0)), totally inside face f (inside face g). By minimality of D we know that the paths $A\#\lambda$ and $B\#\mu$ are non-null-homotopic. We show that, assuming that the C_i are chosen so that under minimality of D, $CC(\lambda\cup\mu)$ is minimum, at least one of the closed curves D_1 and D_2 defined by:

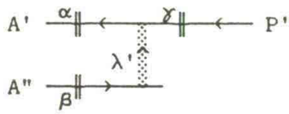
(98) $$\text{path}(D_1) = A\#\lambda\#P^{-1}\#\mu\#B\#P \;;$$
$$\text{path}(D_2) = \lambda\#A\#Q\#B\#\mu\#Q^{-1}$$

is an example contradicting the minimality of D, unless $\overline{P} = (\overline{Q})^{-1}$. We will actually show the last inequality in: $cr(G,D_1) + cr(G,D_2) = 2\cdot cr(G,D) < 2\cdot\Sigma_i mincr(C_i,D) \leq \Sigma_i mincr(C_i,D_1) + \Sigma_i mincr(C_i,D_2)$. Note that if $\overline{P} \neq (\overline{Q})^{-1}$, then each of D_1,D_2 is 'smaller' than D in the sense of (43)(i). Without loss of generality we take $CC(P) \leq CC(Q)$. In our proof we distinguish three cases referring to whether CC(P) and/or CC(Q) are zero or not.

<u>Case 1</u>. $CC(Q) > CC(P) > 0$. We claim that both D_1 and D_2 have liftings passing from one $\gamma \in \Gamma$ to another. To see this consider for instance part $P \# A \# \lambda \# P^{-1}$ of D_1.

In case $CC(A) > 0$, consider lifting λ' of λ, liftings A' and A'' of A, lifting P' of P, and $\alpha, \beta, \gamma \in \Gamma$ so that we have:

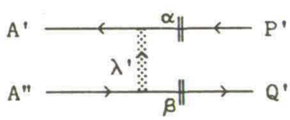
(99)



By the minimality of $CC(\lambda \cup \mu)$ we have that no lifting $\gamma' \in \Gamma$ crosses λ' twice, that $\beta \neq \gamma$, and that none of β and γ cross λ'. This proves that going along P', $(\lambda')^{-1}$ and $(A'')^{-1}$, one passes from one lifting $\gamma' \in \Gamma$ to another.

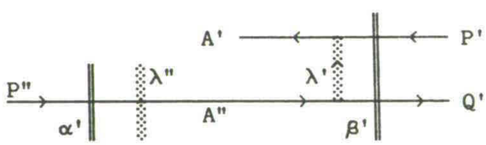
On the other hand, if $CC(A) = 0$, then consider liftings P' of P, A', A'' of A, λ' of λ, Q' of Q, $\alpha, \beta \in \Gamma$ as in:

(100)



Now $\alpha = \beta$, since otherwise the curve along $P \# \lambda^{-1} \# Q \# B$ forms an example contradicting the minimality of D. It follows that, if P' and P'' are liftings of P, λ' and λ'' are liftings of λ and A' and A'' are liftings of A, and we are in the following situation:

(101)

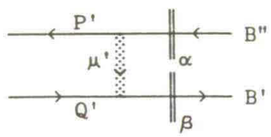


then liftings $\alpha', \beta' \in \Gamma$ crossing P'' and P', respectively, are necessarily distinct. [Here denotes the meeting point of part of a lifting $\delta \in D$ and part of a lifting of λ with one end point on δ, stressing the fact that we cannot tell on which side of δ λ'' is situated.] Hence the part of the lifting of D_1 going along P'', A'', λ' and $(P')^{-1}$ passes from one $\gamma \in \Gamma$ to another (regardless to whether the curve along $A \# \lambda$ flips its orientation or not). Similar results hold fold for part $P^{-1} \# \mu \# B \# P$ of D_1, and for parts $Q^{-1} \# \lambda \# A \# Q$ and $Q \# B \# \mu \# Q^{-1}$ of D_2.

Case 2. $CC(P) = 0$, $CC(Q) = 0$. W.l.o.g. $CC(B) > 0$. We distinguish subcases taking into account whether $CC(A)$ is zero or not.

Subcase 2.1. $CC(A) = 0$. We consider liftings B', B'' of B, P' of P, Q' of Q, μ' of μ and liftings $\alpha, \beta \in \Gamma$ closest to μ' as in:

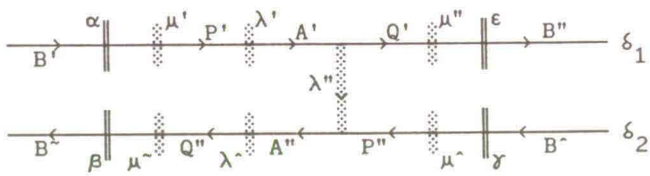
(102)



then $\alpha = \beta$ (since if $\alpha \neq \beta$ then the curve along $B \# \mu$ yields an example contradicting the minimality of D).

Next consider liftings $\delta_1, \delta_2 \in D$ traversing the end points of a lifting λ'' of λ so that we have:

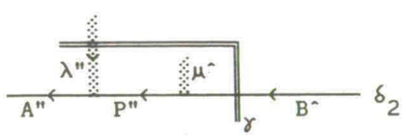
(103)



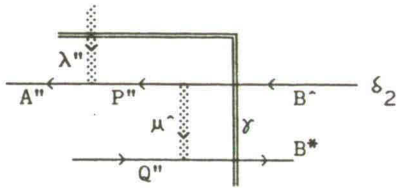
Here $B', B'', B^{\wedge}, B^{\sim}$ denote liftings of B etc.. For liftings $\alpha, \beta, \gamma, \varepsilon \in \Gamma$ we know that $\alpha \neq \varepsilon$ and $\beta \neq \gamma$ as each lifting of D crosses each lifting in Γ at most once. Furthermore we claim that neither γ nor ε crosses λ''.

To see this, suppose for instance that γ crosses λ''. Let $\gamma(x) \in \delta_2$ and $\gamma(y) \in \lambda''$. Then it is easily seen that $|y-x| < 1$. If $\mu^{\wedge}$ is situated as in:

(104)



then re-routing the curve with lifting γ along part of λ and part of D decreases $\Sigma_i cr(D,C_i)$ or $CC(\lambda \cup \mu)$. If $\mu\hat{}$ is situated on the other side of δ_2, then we have:

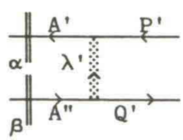
(105) 

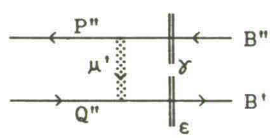
In this case $CC(\lambda \cup \mu)$ is decreased by re-routing γ (and the corresponding closed curve C_i) along λ",P" and $\mu\hat{}$.

Finally we have $\gamma=\varepsilon$, since otherwise the curve D* defined by path(D*) = $Q\#B\#P\#\lambda^{-1}$ would yield an example contradicting the minimality of D.

As $\gamma=\varepsilon$, and as no lifting γ' in Γ intersects λ" more than once, it follows directly that neither α nor β can cross λ". Hence both D_1 and D_2 have liftings passing from one lifting $\gamma' \in \Gamma$ to another.

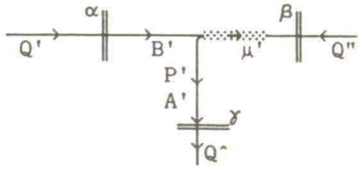
Subcase 2.2. $CC(A) > 0$. Considering liftings λ" of λ, μ' of μ and liftings $\alpha,\beta,\gamma,\varepsilon \in \Gamma$ as in:

(106) 

and 

We have $\alpha=\beta$ or $\gamma=\varepsilon$, since otherwise the curve along $A\#\lambda$ or the curve along $B\#\mu$ forms an example contradicting the minimality of D. So we may assume that $\gamma=\varepsilon$. With arguments similar to those used in subcase 2.1 it follows that both curves D_1,D_2 have liftings passing from one lifting $\gamma' \in \Gamma$ to another.

Case 3. $CC(P) = 0$, $CC(Q) > 0$. Consider a lifting μ' of μ and liftings $\alpha,\beta,\gamma \in \Gamma$ closest to μ'. We have:

(107) 

Lifting α crosses Q' or B', lifting β crosses Q" and lifting γ crosses A' or Q^. We have $\alpha\neq\beta$. Since, otherwise the curve along $B\#\mu$ or the curve along $Q\#\mu^{-1}\#P\#A$ forms an example contradicting the minimality of D.

By the minimality of $CC(\lambda\cup\mu)$, we know that β does not cross μ'.

We claim that lifting α does not cross μ'. For suppose α crosses μ'. Then, by minimality of $CC(\lambda\cup\mu)$, α does not cross B'. Again the curve going along $Q\#\mu^{-1}\#P\#A$ forms an example contradicting the minimality of D.

So we may conclude that each lifting of the path $Q\#B\#\mu\#Q^{-1}$ passes from one lifting $\gamma'\in\Gamma$ to another. The same result holds for $Q^{-1}\#\lambda\#A\#Q$. It follows that if $\overline{P}\neq(\overline{Q})^{-1}$, then D_2 is an example contradicting the minimality of D. This settles the final case in the proof of our claim.

□

This last claim completes the proof of the reduction theorem. □□□

SECTION 4 - APPLICATIONS OF THE REDUCTION THEOREM

In this section we show that theorems 3 and 5 are indeed corollaries of our reduction theorem. We repeat:

Theorem 3. *Let* $G = (V,E)$ *be a graph embedded on a compact orientable surface* S. *Let* $C_1,\ldots,C_k$ *be pairwise disjoint and simple closed curves on* S. *If there exists a closed curve* D *on* S *not intersecting* V *for which:*

$$(108) \qquad \sum_{i=1}^{k} \mathrm{mincr}(C_i,D) \;>\; \mathrm{cr}(G,D)\ ,$$

then there exists one satisfying moreover the 'simplicity conditions' (2).

Proof. Immediate, as the curves C_i are necessarily orientation preserving. □

Next we prove:

Theorem 5. *Let* S *be a compact surface. Let* $G = (V(G),E(G))$ *be a graph embedded on* S. *Let* $C_1,\ldots,C_k: S_1 \rightarrow S$ *be simple pairwise disjoint closed curves on* S, *each of them non-null-homotopic. If there exists a closed curve* $D: S_1 \rightarrow S$, *with the property that*

$$(109) \qquad \mathrm{cr}(G,D) < \Sigma_{i=1}^{k} \mathrm{mincr}(C_i,D),$$

then there exists one satisfying the following further conditions:

(110) (i) D *intersects* G *only in* V(G) ;

(ii) D *has no self-crossings* ;

(iii) *the intersection-sequence* $\overline{D}$ *of* D *is simple or semi-simple* ;

(iv) *if* v_1,v_2 *are vertices of* G, *and* f *is a face of* G, *so that part* D_{12} *of* D *traverses* f, *going from* v_1 *to* v_2, *and part* D_{21} *traverses* f, *going from* v_2 *to* v_1, *then* $D_{12} \sim (D_{21})^{-1}$.

Proof.

The theorem trivially holds if S is equal to the 2-sphere, since then each closed curve is null-homotopic, so $k = 0$.

In case S is the projective plane, it is again easy to see that the theorem holds. In this case any two non-null-homotopic closed curves intersect, hence $k \leq 1$. Therefore a closed curve D satisfying (111) necessarily has $cr(G,D) = 0$ and $mincr(C_1,D) = 1$. We claim that we may assume that D is simple. To see this, first note that we may assume from the tameness of the surface that $cr(D) < \infty$. Next, it is obvious that any split of D into $D \circ \varphi = D_1 \# D_2$ yields a non-null-homotopic closed curve D_1 or D_2 with $cr(D_i) < cr(D)$, $cr(G,D_i) = 0$ and $mincr(C_1,D_i) = 1$. Repeating this we obtain a curve with the required properties.

So from now on, let S and G be fixed, where S is a compact surface, not equal to the 2-sphere or the projective plane, and $G = (V(G),E(G))$ is a graph embedded on S. G is identified with its embedding on S. Let $C_1,\dots,C_k:S_1 \to S$ be closed curves on S, pairwise disjoint, each of them simple and non-null-homotopic.

Now suppose there is a closed curve $D:S_1 \to S$ satisfying

$$(111) \qquad cr(G,D) < \Sigma_{i=1}^{k} mincr(C_i,D) .$$

Claim 1. We may assume that D is not homotopic to C_i^q for some i and some $q \in \mathbb{Z}$.

Proof of Claim 1. Suppose $D \sim C_i^q$ for some i,q. Then $mincr(C_j,D) = mincr(C_j,C_i^q) = 0$, if $j \neq i$, and ≤ 1, if $j=i$. So $cr(G,D) = 0$, q is odd, and D and C_i are orientation reversing. By lemma (30) we can split off from D a simple curve $D' \sim C_i$ with $cr(G,D') = 0$. Thus (110) is satified. □

This implies:

Claim 2. We may assume:

$$(112) \qquad mincr(C_i^2,D) = 2 \cdot mincr(C_i,D).$$

Proof of Claim 2. To see that we can do this, observe that if $\tilde{C} \sim C_i$ and $\tilde{D} \sim D$ attain $cr(\tilde{C},\tilde{D}) = mincr(C_i,D)$, with $\tilde{C}$ simple, then each lifting δ of $\tilde{D}$ to S' intersects each lifting γ of $\tilde{C}$ to S' at most once. Then a small enough perturbation of $\tilde{C}_i^2$ yields a curve $C' \sim C_i^2$ such that each lifting δ'' of $\tilde{D}$ to S' intersects each lifting γ'' of C' to S' at most once. As a result $mincr(C_i^2,D) = cr(C',\tilde{D}) = 2\cdot cr(\tilde{C},\tilde{D}) = 2\cdot mincr(C_i,D)$. □

Define $\bar{G}$ to be the graph resulting from 'blowing up the vertices of G on S until they touch'. $\bar{G}$ is described in section 2. Let $\Psi:S\backslash V(\bar{G})\rightarrow S$ denote a continuous transformation, so that $\Psi(disc(v)) = v$, for each vertex v of G, and so that $\Psi(f') = f$, and Ψ restricted to f' is a homeomorphism, for each face f' of $\bar{G}$ corresponding to face f of G.

We have, without loss of generality:

(113) D does not intersect $V(\bar{G})$ and $cr(\bar{G},D) = 2\cdot cr(G,D)$.

We now introduce a new set of simple and pairwise disjoint closed curves $C'_{i'}$, for $i' = 1,2,\ldots,k'$, by 'duplication' of the C_i. This is done as follows: if C_i is orientation-preserving, then $C'_{i'}$ and $C'_{i''}$ arise from C_i by 'drawing' two simple and pairwise disjoint curves 'close to' C_i, one to the left and one to the right of C_i. If C_i is orientation-reversing, then one curve $C'_{i'}$ arises by drawing a simple closed curve close to C_i, going around twice along C_i. The resulting curves C'_i are orientation-preserving and we have:

(114) $cr(\bar{G},D) < \Sigma_{i'} mincr(C'_{i'},D)$,

as $cr(\bar{G},D) = 2\cdot cr(G,D) < 2\cdot\Sigma_i mincr(C_i,D) = \Sigma_{i'} mincr(C'_{i'},D)$. Application of the reduction theorem now implies the existence of a curve D not passing $V(\bar{G})$ and satisfying $cr(\bar{G},D) < \Sigma_{i'} mincr(C'_{i'},D)$, and satisfying the additional conditions (110) (with respect to graph $\bar{G}$ and curves C'_i). It is easily seen that the image of $\Psi(D)$ contains as a subset the image of a curve satisfying required conditions. □□□

REFERENCES

[1] L.E.J. Brouwer, Über eindeutige stetige Transformationen von Flächen in sich, *Math. Ann.* **69** (1910) 176-180.

[2] M. Dehn and P. Heegaard, Analysis situs, in: *Encyklopädie der mathematischen Wissenschaften, mit Einschluss ihrer Anwendungen, III. Band: Geometrie, I. Teil, I. Hälfte* (W.Fr. Meyer and H. Mohrmann, eds.), Teubner, Leipzig, 1907-1910, pp. 153-220.

[3] C.A.J. Hurkens, On the Existence of an Integral Potential in a Weighted Bidirected Graph, *Linear Algebra and its Applications*, **114/115** (1989) 541-553.

[4] C.A.J. Hurkens, Reduction of Parity-Cut-Conditions on Compact Surfaces, to appear.

[5] B. von Kerékjártó, *Vorlesungen über Topologie I: Flächentopologie*, Springer-Verlag, Berlin, 1923.

[6] H. Poincaré, 5e complément à l'Analysis Situs, *Rend. circ. mat. Palermo*, **18**, (1904) 45-110.

[7] A. Schrijver, Decomposition of Graphs on Surfaces and a Homotopic Circulation Theorem, to appear in: *Journal of Combinatorial Theory (B)*.

[8] A. Schrijver, Disjoint Circuits of Prescribed Homotopies in a Graph on a Compact Surface, to appear in: *Journal of Combinatorial Theory (B)*.

SAMENVATTING.

Nieuwe Technieken in de Meetkundige en Discrete Optimalisering

Dit proefschrift is een collectie van zeven artikelen over problemen van meetkundig en discreet karakter. Behandeld worden vragen die opduiken bij de bestudering van problemen in de combinatorische optimalisering, de geheeltallige programmering en de discrete meetkunde. In de onderhavige artikelen ligt de nadruk vooral op het scherp schatten van (boven)grenzen en het maximaal afzwakken (ook wel 'uitdunnen' genoemd) van bepaalde nodige en voldoende voorwaarden.

Vier van de artikelen (nummers 1,2,4 en 6) zijn reeds verschenen in wetenschappelijke tijdschriften en twee (nummers 3 en 5) zijn geaccepteerd voor publicatie. Het laatste en meest recente artikel is nog niet ter publicatie aangeboden.

Het eerste artikel, **'On fractional multicommodity flows and distance functions'**, is een gezamenlijk werk met A. Schrijver en É. Tardos. Hierin worden enkele resultaten besproken die betrekking hebben op voorwaarden voor het bestaan van (geheeltallige) multicommodity flows in planaire grafen. Onder meer wordt onderzocht onder welke condities het bestaan van een fraktionele oplossing van het probleem de existentie van een geheeltallige oplossing impliceert.

Het tweede artikel, **'On the size of systems of sets every *t* of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems'**, is opnieuw gezamenlijk werk met A. Schrijver.

In dit werk geven we scherpe grenzen voor de relatieve grootte van een collectie verzamelingen met de eigenschap dat elk element in hooguit k verzamelingen voorkomt en dat elk t-tal verzamelingen een zogenaamd 'systeem van onderscheiden representanten' heeft (System of Distinct Representatives). In het bewijs dat de grenzen scherp zijn maken we gebruik van resultaten uit de extremale grafentheorie.

Het puur combinatorische resultaat is toe te passen bij het afschatten van het zogenaamde 'worst-case' gedrag van bepaalde heuristieken voor het vinden van een maximale deelcollectie paarsgewijs disjunkte verzamelingen uit een gegeven collectie verzamelingen ter grootte k.

Het onderzoek dat leidde tot bovengenoemd resultaat vond zijn oorsprong in de studie naar het gedrag van zekere heuristieken voor packing-problemen voorgesteld door J.K. Lenstra.

Het derde en het vierde artikel zijn sterk gerelateerd. Het derde artikel is genaamd **'On the diameter of the edge cover polytope'**.

Het *edge cover polytoop* van een graaf is het convex omhulsel van de karakteristieke vectoren van de edge covers van de graaf. Wanneer een algoritme, zoals bijvoorbeeld dat welke gebruikt wordt bij de simplex methode, bij bepaalde iteratieslagen via de ribben van het polytoop van hoekpunt naar hoekpunt verhuist, dan is het aantal 'verhuizingen' een maat voor de complexiteit van het algoritme. Noemen we het minimaal aantal verhuizingen om van het ene hoekpunt in het andere terecht te komen de *afstand* tussen deze hoekpunten, dan is de *diameter* van het polytoop, zijnde de maximale afstand tussen enig tweetal hoekpunten, ook een maat voor de complexiteit van het algoritme.

In het artikel wordt aangegeven wanneer twee hoekpunten van het edge cover polytoop van een graaf aangrenzend zijn. Voorts wordt een bovengrens geformuleerd voor de afstand tussen twee hoekpunten en hiermee is het dan mogelijk de diameter van het polytoop expliciet te berekenen.

Bovengenoemde resultaten konden worden gegeneraliseerd. Het complement van een edge cover van een graaf G is namelijk een b-matching in G, met $b(v) = \deg_G(v)-1$, voor elk punt v. Deze generalisatie wordt beschreven in het vierde artikel **'On the diameter of the *b*-matching polytope'**. Het hoofdresultaat hierin luidt als volgt: de diameter van het *b*-matching polytoop is gelijk aan de cardinaliteit van de grootste *b*-matching.

In het vijfde artikel, **'Blowing up Convex Sets in the Plane'**, wordt een probleem opgelost dat gesteld werd door R. Kannan en L. Lovász. Dit probleem in de Euclidische meetkunde kwamen zij tegen in hun studie van de theorie van lattices en de meetkunde der getallen. Het resultaat dat wordt beschreven in het artikel luidt als volgt.

Voor een convexe verzameling K in $\mathbb{R}^2$ en een lattice $\mathcal{L}$, met de eigenschap dat elke lijn in $\mathbb{R}^2$ niet-lege doorsnede heeft met $K + \mathcal{L}$, geldt dat $\alpha \cdot K + \mathcal{L}$ $\mathbb{R}^2$ overdekt, als $\alpha \geq 1 + 2/\sqrt{3}$. Deze grens is scherp.

De twee laatste artikelen vertonen grote samenhang. Beide gaan uit van de oplossing van A. Schrijver voor het probleem van het vinden van punt-disjunkte circuits van voorgeschreven homotopie in een graaf ingebed op een compact oppervlak. Problemen van deze aard vindt onder meer bij het ontwerp van VLSI-chips. Daarnaast speelt dit onderwerp een rol in het grote Graph Minors projekt van N. Robertson en P.D. Seymour.

Een oude versie van het algoritme van Schrijver bevat als onderdeel een subroutine voor het oplossen van een systeem van lineaire ongelijkheden in integers. In het algemeen gesproken is dit, in de zin van de complexiteitstheorie, een 'moeilijk' probleem. Als gevolg van de speciale structuur van de technologie-matrix van het onderhavige probleem is de opgave echter handelbaar. Beschouwen we de matrix als verbindingsmatrix van een zogenaamde 'bidirected' graaf, dan kunnen nodige en voldoende voorwaarden opgesteld worden die het bestaan van een geheeltallige oplossing garanderen. Deze condities worden geformuleerd in termen van voorwaarden op cykels in de graaf.

In het zesde artikel, **'On the existence of an integral potential in a weighted bidirected graph'**, wordt aangetoond dat het voldoende is de bovengenoemde condities op te leggen aan een beperkte verzameling van redelijk simpele cykels in de graaf.

In Schrijvers stelling worden nodige en voldoende voorwaarden gegeven voor het bestaan van punt-disjunkte circuits van voorgeschreven homotopie in een graaf ingebed op een compact oppervlak. In het zevende artikel, **'Reduction of cut-conditions on compact surfaces'**, geven we een gedeeltelijke reductie van deze verzameling van nodige en voldoende voorwaarden. Dit artikel vormt een eerste stap in de transformatie van Schrijvers stelling naar een zogenaamde 'goede karakterisering'.

Dankbetuiging.

De in dit proefschrift vermelde resultaten vormen een deel van de oogst van een vruchtbare periode van onderzoek die aanving per 1 september 1985 toen ondergetekende als onderzoekmedewerker in dienst trad bij Z.W.O., de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek. Via het Mathematisch Centrum ondersteunde Z.W.O. (later N.W.O.) het projekt "Polyhedrale en Polynomiale Methoden in de Combinatorische Optimalisering" dat onder leiding stond van prof. dr. Alexander Schrijver. Onder diens bezielende begeleiding heb ik veel geleerd over voor mij nieuwe gebieden uit de wiskunde, over het vernieuwen van de wiskunde en het verleggen van bekende grenzen, over presentatie in woord en geschrift. Ik heb niet alleen veel technische kennis opgedaan, maar ook de lol in de wiskunde en de uitdaging die uitgaat van vele mathematische mysteries zijn blijven bestaan. Ik ben Lex hiervoor uiterst dankbaar en hoop dan ook dat wij nog veel mogen samenwerken.

Daarnaast wil ik iedereen bedanken die mij het wetenschappelijke pad gekruist is. Met name gaat mijn dank uit naar Bert Gerards, Willem Haemers en ook Chris Wildhagen, die samen met Lex vaak mijn half afgemaakte verhalen moesten aanhoren. Met hun geduld en kritiek kwam ik steeds weer verder.

Cor Hurkens, augustus 1989